PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : A61K 39/00, C12Q 1/68, A01N 43/04 — A1

(11) International Publication Number: WO 98/24475

(43) International Publication Date: 11 June 1998 (11.06.98)

(21) International Application Number: PCT/US97/22104

(22) International Filing Date: 5 December 1997 (05.12.97)

(30) Priority Data:

08/759,625	5 December 1996 (05.12.96)	US
08/823,745	25 March 1997 (25.03.97)	US
08/891,928	14 July 1997 (14.07.97)	US

(71) Applicant *(for all designated States except US)*: ASTRA AKTIEBOLAG [SE/SE]; S-151 85 Södertälje (SE).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: SMITH, Douglas [US/US]; 2 Mayflower Lane, Gloucester, MA 01930 (US). ALM, Richard, A. [AU/US]; 28 Russett Hill Road, Ashland, MA 01721 (US). DOIG, Peter, C. [CA/US]; 14 Notre Dame Road, Acton, MA 01720 (US). KABOK, Zita [US/US]; 19 University Drive, Natick, MA 01760 (US). CASTRIOTTA, Lillian, Marie [US/US]; 119 Charlesbank Road, Newton, MA 02158 (US).

(74) Agents: MANDRAGOURAS, Amy, E. et al.; Lahive & Cockfield, LLP, 60 State Street, Boston, MA 02109 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published

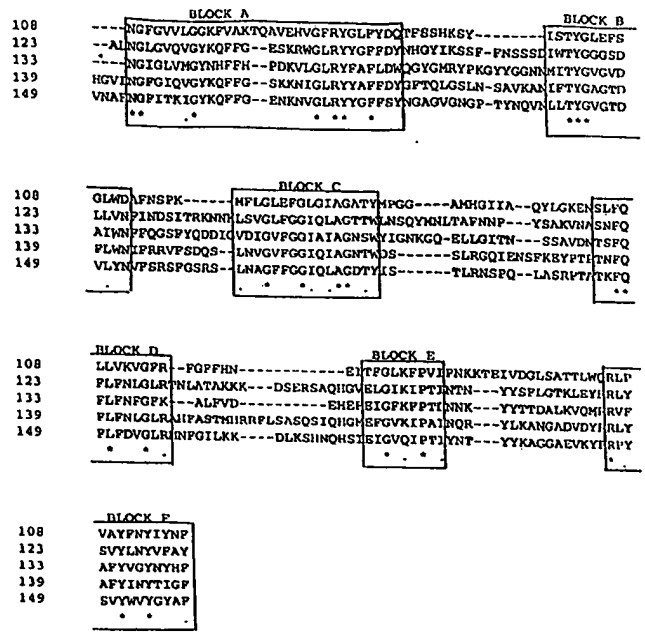
With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO *HELICOBACTER PYLORI* AND VACCINE COMPOSITIONS THEREOF

(57) Abstract

Recombinant or substantially pure preparations of *H. pylori* polypeptides are described. The nucleic acids encoding the polypeptides also are described. The *H. pylori* polypeptides are useful for diagnostics and vaccine compositions, wherein the figure depicts an amino acid sequence alignment of five *H. pylori* proteins.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Rômania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO *HELICOBACTER PYLORI* AND VACCINE COMPOSITIONS THEREOF

Background of the Invention

Helicobacter pylori is a gram-negative, S-shaped, microaerophilic bacterium that was discovered and cultured from a human gastric biopsy specimen. (Warren, J.R. and B. Marshall, (1983) *Lancet* 1: 1273-1275; and Marshall et al., (1984) *Microbios Lett.* 25: 83-88). *H. pylori* has been strongly linked to chronic gastritis and duodenal ulcer disease. (Rathbone et. al., (1986) *Gut* 27: 635-641). Moreover, evidence is accumulating for an etiologic role of *H. pylori* in nonulcer dyspepsia, gastric ulcer disease, and gastric adenocarcinoma. (Blaser M. J., (1993) *Trends Microbiol.* 1: 255-260). Transmission of the bacteria occurs via the oral route, and the risk of infection increases with age. (Taylor, D.N. and M. J. Blaser, (1991) *Epidemiol. Rev* 13: 42-50). *H. pylori* colonizes the human gastric mucosa, establishing an infection that usually persists for decades. Infection by *H. pylori* is prevalent worldwide. Developed countries have infection rates over 50% of the adult population, while developing countries have infection rates reaching 90% of the adults over the age of 20. (Hopkins R. J. and J. G. Morris (1994) *Am. J. Med.* 97: 265-277).

The bacterial factors necessary for colonization of the gastric environment, and for virulence of this pathogen, are poorly understood. Examples of the putative virulence factors include the following: urease, an enzyme that may play a role in neutralizing gastric acid pH (Eaton et al., (1991) *Infect. Immunol.* 59: 2470-2475; Ferrero, R.L. and A. Lee (1991) *Microb. Ecol. Hlth. Dis.* 4: 121-134; Labigne et al., (1991) *J. Bacteriol.* 173: 1920-1931); the bacterial flagellar proteins responsible for motility across the mucous layer. (Hazell et al., (1986) *J. Inf. Dis.* 153: 658-663; Leying et al., (1992) *Mol. Microbiol.* 6: 2863-2874; and Haas et al., (1993) *Mol. Microbiol.* 8: 753-760); Vac A, a bacterial toxin that induces the formation of intracellular vacuoles in epithelial cells (Schmitt, W. and R. Haas, (1994) *Molecular Microbiol.* 12(2): 307-319); and several gastric tissue-specific adhesins. (Boren et al., (1993) *Science* 262: 1892-1895; Evans et al., (1993) *J. Bacteriol.* 175: 674-683; and Falk et al., (1993) *Proc. Natl. Acad. Sci.* USA 90: 2035-203).

Numerous therapeutic agents are currently available that eradicate *H. pylori* infections *in vitro*. (Huesca et. al., (1993) *Zbl. Bakt.* 280: 244-252; Hopkins, R. J. and J. G. Morris, supra). However, many of these treatments are suboptimally effective *in vivo* because of bacterial resistance, altered drug distribution, patient non-compliance or poor drug availabilty. (Hopkins, R. J. and J. G. Morris, supra). Treatment with antibiotics combined with bismuth are part of the standard regime used to treat *H. pylori* infection.

(Malfertheiner, P. and J. E. Dominguez-Munoz (1993) *Clinical Therapeutics* 15 Supp. B: 37-48). Recently, combinations of a proton pump inhibitors and a single antibiotic have been shown to ameliorate duodenal ulcer disease. (Malfertheiner, P. and J. E. Dominguez-Munoz supra). However, methods employing antibiotic agents can have the problem of the emergence of bacterial strains which are resistant to these agents. (Hopkins, R. J. and J. G. Morris, supra). These limitations demonstrate that new more effective methods are needed to combat *H. pylori* infections *in vivo*. In particular, the design of new vaccines that may prevent infection by this bacterium is highly desirable.

Summary of the Invention

This invention relates to novel genes, e.g., genes encoding polypeptides such as bacterial surface proteins, from the organism *Helicobacter pylori* (*H. pylori*), and other related genes, their products, and uses thereof. The nucleic acids and peptides of the present invention have utility for diagnostic and therapeutics for *H. pylori* and other *Helicobacter* species. They can also be used to detect the presence of *H. pylori* and other *Helicobacter* species in a sample; and for use in screening compounds for the ability to interfere with the *H. pylori* life cycle or to inhibit *H. pylori* infection. More specifically, this invention features compositions of nucleic acids corresponding to entire coding sequences of *H. pylori* proteins, including surface or secreted proteins or parts thereof, nucleic acids capable of binding mRNA from *H. pylori* proteins to block protein translation, and methods for producing *H. pylori* proteins or parts thereof using peptide synthesis and recombinant DNA techniques. This invention also features antibodies and nucleic acids useful as probes to detect *H. pylori* infection. In addition, vaccine compositions and methods for the protection or treatment of infection by *H. pylori* are within the scope of this invention.

Detailed Description of the Drawings

Figure 1 depicts an amino acid sequence alignment of five *H. pylori* proteins (depicted in the single letter amino acid code and designated by their amino acid Sequence ID Numbers; shown N-terminal to C-terminal, left to right).

Figure 2 depicts the N-terminal portion of three *H. pylori* proteins (depicted in the single letter amino acid code and designated by their amino acid Sequence ID Numbers; shown N-terminal to C-terminal, left to right).

Detailed Description of the Invention

In one aspect, the invention features a recombinant or substantially pure preparation of *H. pylori* polypeptide of SEQ ID NO: 98. The invention also includes substantially pure nucleic acid encoding an *H. pylori* polypeptide of SEQ ID NO: 98, such nucleic acid is contained in SEQ ID NO: 1. The *H. pylori* polypeptide sequences of the invention described herein are contained in the Sequence Listing, and the nucleic acids encoding *H. pylori* polypeptides of the invention are contained in the Sequence Listing.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 99, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 2.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 100, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 3.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 101, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 4.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 102, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 5.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 103, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 6.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 104, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 7.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 105, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 8.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 106, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 9.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 107, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 10.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 108, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 11.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 109, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 12.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 110, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 13.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 111, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 14.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 112, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 15.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 113, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 16.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 114, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 17.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 115, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 18.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 116, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 19.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 117, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 20.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 118, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 21.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 119, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 22.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 120, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 23.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 121, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 24.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 122, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 25.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 123, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 26.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 124, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 27.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 125, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 28.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 126, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 29.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 127, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 30.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 128, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 31.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 129, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 32.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 130, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 33.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 131, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 34.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 132, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 35.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 133, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 36.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 134, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 37.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 135, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 38.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 136, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 39.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 137, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 40.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 138, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 41.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 139, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 42.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 140, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO:43.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 141, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 44.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 142, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 45.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 143, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 46.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 144, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 47.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 145, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 48.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 146, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 49.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 147, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 50.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 148, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 51.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 149, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 52.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 150, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 53.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 151, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 54.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 152, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 55.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 153, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 56.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 154, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 57.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 155, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 58.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 156, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 59.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 157, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 60.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 158, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 61.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 159, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 62.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 160, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 63.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 161, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 64.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 162, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 65.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 163, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 66.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 164, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 67.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 165, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 68.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 166, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 69.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 167, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 70.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 168, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 71.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 169, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 72.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 170, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 73.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 171, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 74.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 172, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 75.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 173, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 76.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 174, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 77.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 175, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 78.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 176, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 79.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 177, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 80.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 178, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 81.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 179, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 82.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 180, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 83.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 181, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 84.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 182, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 85.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 183, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 86.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 184, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 87.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 185, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 88.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 186, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 89.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 187, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 90.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 188, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 91.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 189, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 92.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 190, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 93.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 191, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 94.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 192, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 95.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 193, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 96.

In another aspect, the invention features a substantially pure nucleic acid encoding an *H. pylori* polypeptide having an amino acid sequence of SEQ ID NO: 194, such as a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 97.

In another aspect, the invention features an isolated nucleic acid having a nucleotide sequence encoding an *H. pylori* polypeptide at least about 60% homologous to an amino acid sequence selected from the group consisting of SEQ ID NO: 98-SEQ ID NO: 194. In a preferred embodiment, the isolated nucleic acid includes a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 97, or a complement thereof.

In another aspect, the invention features an isolated nucleic acid having a nucleotide sequence encoding an *H. pylori* polypeptide selected from the group consisting of SEQ ID NO: 98-SEQ ID NO: 194.

In another aspect, the invention features an isolated nucleic acid which encodes an *H. pylori* polypeptide, having a nucleotide sequence at least about 60% homologous to a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 97, or a complement thereof.

In another aspect, the invention features an isolated nucleic acid molecule encoding an *H. pylori* polypeptide, having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 97, or a complement thereof.

In another aspect, the invention features an isolated nucleic acid having a nucleotide sequence of at least 8 nucleotides in length, wherein the sequence hybridizes under stringent hybridization conditions to a nucleic acid having a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 97, or a complement thereof.

Particularly preferred is an isolated nucleic acid having a nucleotide sequence encoding an *H. pylori* cell envelope polypeptide or a fragment thereof, the nucleic acid selected from the group consisting of SEQ ID NO: 63, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 61,

SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 91, SEQ ID NO: 94, SEQ ID NO: 5, SEQ ID NO: 11, SEQ ID NO: 26, SEQ ID NO: 36, SEQ ID NO: 42, SEQ ID NO: 52, SEQ ID NO: 22, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 6, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 60, SEQ ID NO: 69, and SEQ ID NO: 83, or a complement thereof.

In one embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* flagella-associated polypeptide or a fragment thereof encoded by a nucleic acid having a nucleotide sequence of SEQ ID NO: 63, or a complement thereof.

In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* inner membrane polypeptide or a fragment thereof encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 38, and SEQ ID NO: 39, or a complement thereof.

In another embodiment, the *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in transport encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 43, and SEQ ID NO: 44, or a complement thereof.

In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* outer membrane polypeptide or a fragment thereof encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 61, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 91, SEQ ID NO: 94, SEQ ID NO: 5, SEQ ID NO: 11, SEQ ID NO: 26, SEQ ID NO: 36, SEQ ID NO: 42, SEQ ID NO: 52, SEQ ID NO: 22, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 65, and SEQ ID NO: 66, or a complement thereof.

In another embodiment, the *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue or a fragment thereof encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 36, SEQ ID NO: 42, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID

NO: 52, SEQ ID NO: 61, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 91, and SEQ ID NO: 94, or a complement thereof.

In another embodiment, *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue and a C-terminal tyrosine cluster or a fragment thereof encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 11, SEQ ID NO: 26, SEQ ID NO: 36, SEQ ID NO: 42, and SEQ ID NO: 52, or a complement thereof.

Particularly preferred is an isolated nucleic acid having a nucleotide sequence encoding an *H. pylori* cell envelope polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 160, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 158, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 188, SEQ ID NO: 191, SEQ ID NO: 102, SEQ ID NO: 108, SEQ ID NO: 123, SEQ ID NO: 133, SEQ ID NO: 139, SEQ ID NO: 149, SEQ ID NO: 119, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 162, SEQ ID NO: 163, SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 103, SEQ ID NO: 131, SEQ ID NO: 132, SEQ ID NO: 157, SEQ ID NO: 166, and SEQ ID NO: 180.

In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* flagella-associated polypeptide or a fragment thereof having an amino acid sequence of SEQ ID NO: 160.

In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* inner membrane polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 135, and SEQ ID NO: 136.

In another embodiment, the *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in transport selected from the group consisting of SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 140, and SEQ ID NO: 141.

In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* outer membrane polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 158, SEQ ID NO: 176,

SEQ ID NO: 177, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 188, SEQ ID NO: 191, SEQ ID NO: 102, SEQ ID NO: 108, SEQ ID NO: 123, SEQ ID NO: 133, SEQ ID NO: 139, SEQ ID NO: 149, SEQ ID NO: 119, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 162, and SEQ ID NO: 163.

In another embodiment, the *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue or a fragment thereof selected from the group consisting of SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 108, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 123, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 133, SEQ ID NO: 139, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 149, SEQ ID NO: 158, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 188, and SEQ ID NO: 191.

In another embodiment, the *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue and a C-terminal tyrosine cluster or a fragment thereof selected from the group consisting of SEQ ID NO: 108, SEQ ID NO: 123, SEQ ID NO: 133, SEQ ID NO: 139, and SEQ ID NO: 149.

Particularly preferred is an isolated nucleic acid having a nucleotide sequence encoding an *H. pylori* cytoplasmic polypeptide or a fragment thereof, wherein the nucleic acid is selected from the group consisting of SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 92, and SEQ ID NO: 93, or a complement thereof.

In one embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in mRNA translation, wherein the nucleic acid is selected from the group consisting of SEQ ID NO: 57 and SEQ ID NO: 58, or a complement thereof.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in genome replication, transcription, recombination and repair, wherein the nucleic acid is selected from the group consisting of SEQ ID NO: 86, SEQ ID NO: 87, or a complement thereof.

Particularly preferred is an isolated nucleic acid having a nucleotide sequence encoding an *H. pylori* cytoplasmic polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 154, SEQ ID NO: 155, SEQ ID NO: 183, SEQ ID NO: 184, SEQ ID NO: 185, SEQ ID NO: 186, SEQ ID NO: 189, and SEQ ID NO: 190.

In one embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in mRNA translation selected from the group consisting of SEQ ID NO: 154 and SEQ ID NO: 155.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in genome replication, transcription, recombination and repair selected from the group consisting of SEQ ID NO: 183 and SEQ ID NO: 184.

Particularly preferred is an isolated nucleic acid having a nucleotide sequence encoding an *H. pylori* secreted polypeptide or a fragment thereof, the nucleic acid selected from the group consisting of SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 20, SEQ ID NO: 25, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 53 SEQ ID NO: 64, SEQ ID NO: 67, SEQ ID NO: 70, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 81, SEQ ID NO: 82, SEQ ID NO: 90, SEQ ID NO: 95, and SEQ ID NO: 97, or a complement thereof.

Particularly preferred is an isolated nucleic acid having a nucleotide sequence encoding an *H. pylori* secreted polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 107, SEQ ID NO: 109, SEQ ID NO: 117, SEQ ID NO: 122, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 142, SEQ ID NO: 143, SEQ ID NO: 150 SEQ ID NO: 161, SEQ ID NO: 164, SEQ ID NO: 167, SEQ ID NO: 174, SEQ ID NO: 175, SEQ ID NO: 178, SEQ ID NO: 179, SEQ ID NO: 187, SEQ ID NO: 192, and SEQ ID NO: 194.

Particularly preferred is an isolated nucleic acid having a nucleotide sequence encoding an *H. pylori* cellular polypeptide or a fragment thereof, the nucleic acid selected from the group consisting of SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 21, SEQ ID NO: 33, SEQ ID NO: 37, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 47, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56 SEQ ID NO: 59, SEQ ID NO: 62, SEQ ID NO: 68, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, and SEQ ID NO: 96, or a complement thereof.

Particularly preferred is an isolated nucleic acid having a nucleotide sequence encoding an *H. pylori* cellular polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 118, SEQ ID NO: 130, SEQ ID NO: 134, SEQ ID NO: 137, SEQ ID NO: 138, SEQ ID NO: 144, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 153, SEQ ID NO: 156, SEQ ID NO: 159, SEQ ID NO: 165, SEQ ID NO: 168, SEQ ID NO: 169, SEQ ID NO: 170, SEQ ID NO: 171, SEQ ID NO: 172, SEQ ID NO: 173, and SEQ ID NO: 193.

In another aspect, the invention features a probe having a nucleotide sequence consisting of at least 8 nucleotides of a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 97, or a complement thereof.

In another aspect, the invention features an isolated *H. pylori* polypeptide having an amino acid sequence at least about 60% homologous to an *H. pylori* polypeptide selected from the group consisting of SEQ ID NO: 98-SEQ ID NO: 194.

In another aspect, the invention features an isolated *H. pylori* polypeptide which is encoded by a nucleic acid having a nucleotide sequence at least about 60% homologous to a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 97. In one embodiment, the isolated *H. pylori* polypeptide is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 97.

In another aspect, the invention features an isolated *H. pylori* polypeptide which is encoded by a nucleic acid which hybridizes under stringent hybridization conditions to a nucleic acid selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 97, or a complement thereof.

In another aspect, the invention features an isolated *H. pylori* polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO: 97-SEQ ID NO: 194.

Particularly preferred is an isolated *H. pylori* cell envelope polypeptide or a fragment thereof, wherein the polypeptide is selected from the group consisting of SEQ ID NO: 160, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 158, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 188, SEQ ID NO: 191, SEQ ID NO: 102, SEQ ID NO: 108, SEQ ID NO: 123, SEQ ID NO: 133, SEQ ID NO: 139, SEQ ID NO: 149, SEQ ID NO: 119, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 162, SEQ ID NO: 163, SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 103, SEQ ID NO: 131, SEQ ID NO: 132, SEQ ID NO: 157, SEQ ID NO: 166, and SEQ ID NO: 180.

In one embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* flagella-associated polypeptide or a fragment thereof having an amino acid sequence of SEQ ID NO: 160.

In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* inner membrane polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 135, and SEQ ID NO: 136.

In another embodiment, the *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in transport selected from the group consisting of SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 135, and SEQ ID NO: 136.

In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* outer membrane polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 158, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 188, SEQ ID NO: 191, SEQ ID NO: 102, SEQ ID NO: 108, SEQ ID NO: 123, SEQ ID NO: 133, SEQ ID NO: 139, SEQ ID NO: 149, SEQ ID NO: 119, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 162, and SEQ ID NO: 163.

In another embodiment, the *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue or a fragment thereof selected from the group consisting of SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 108, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 123, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 133, SEQ ID NO: 139, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 149, SEQ ID NO: 158, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 188, and SEQ ID NO: 191.

In another embodiment, the *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue and a C-terminal tyrosine cluster or a fragment thereof selected from the group consisting of SEQ ID NO: 108, SEQ ID NO: 123, SEQ ID NO: 133, SEQ ID NO: 139, and SEQ ID NO: 149.

Particularly preferred is an isolated *H. pylori* cell envelope polypeptide or a fragment thereof, wherein the polypeptide is encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 63, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 61, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 91, SEQ ID NO: 94, SEQ ID NO: 5, SEQ ID NO: 11, SEQ ID NO: 26, SEQ ID NO: 36, SEQ ID NO: 42, SEQ ID NO: 52, SEQ ID NO: 22, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 38, SEQ ID NO: 39,

SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 6, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 60, and SEQ ID NO: 69, SEQ ID NO: 83.

In one embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* flagella-associated polypeptide or a fragment thereof encoded by a nucleic acid having a nucleotide sequence of SEQ ID NO: 63.

In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* inner membrane polypeptide or a fragment thereof encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 38, and SEQ ID NO: 39.

In another embodiment, the *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in transport encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 43, and SEQ ID NO: 44.

In another embodiment, the *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* outer membrane polypeptide or a fragment thereof encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 61, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 91, SEQ ID NO: 94, SEQ ID NO: 5, SEQ ID NO: 11, SEQ ID NO: 26, SEQ ID NO: 36, SEQ ID NO: 42, SEQ ID NO: 52, SEQ ID NO: 22, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 65, and SEQ ID NO: 66.

In another embodiment, the *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue or a fragment thereof encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 36, SEQ ID NO: 42, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 61, SEQ ID NO. 79, SEQ ID NO: 80, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 91, and SEQ ID NO: 94.

In another embodiment, the *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue and a C-terminal tyrosine cluster or a fragment thereof encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 11, SEQ ID NO: 26, SEQ ID NO: 36, SEQ ID NO: 42, and SEQ ID NO: 52.

Particularly preferred is an isolated *H. pylori* cytoplasmic polypeptide or a fragment thereof, wherein the polypeptide is selected from the group consisting of SEQ ID NO: 154, SEQ ID NO: 155, SEQ ID NO: 183, SEQ ID NO: 184, SEQ ID NO: 185, SEQ ID NO: 186, SEQ ID NO: 189, and SEQ ID NO: 190.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in mRNA translation selected from the group consisting of SEQ ID NO: 154 and SEQ ID NO: 155.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in genome replication, transcription, recombination and repair selected from the group consisting of SEQ ID NO: 183 and SEQ ID NO: 184.

Particularly preferred is an isolated *H. pylori* cytoplasmic polypeptide or a fragment thereof, wherein the polypeptide is encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 92, and SEQ ID NO: 93.

In one embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in mRNA translation, wherein the polypeptide is encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 57, and SEQ ID NO: 58.

In another embodiment, the *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in genome replication, transcription, recombination and repair, wherein the polypeptide is encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 86 and SEQ ID NO: 87.

Particularly preferred is an isolated *H. pylori* cellular polypeptide or a fragment thereof, wherein the polypeptide is selected from the group consisting of SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 118, SEQ ID NO: 130, SEQ ID NO: 134, SEQ ID NO: 137, SEQ ID NO: 138, SEQ ID NO: 144, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 153, SEQ ID NO: 156, SEQ ID NO: 159, SEQ ID NO: 165, SEQ ID NO: 168, SEQ ID NO: 169, SEQ ID NO: 170, SEQ ID NO: 171, SEQ ID NO: 172, SEQ ID NO: 173, and SEQ ID NO: 193.

Particularly preferred is an isolated *H. pylori* cellular polypeptide or a fragment thereof, wherein the polypeptide is encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 21, SEQ ID NO: 33, SEQ ID NO: 37, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 47, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56 SEQ ID NO: 59, SEQ ID NO: 62, SEQ ID NO: 68, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, and SEQ ID NO: 96.

Particularly preferred is an isolated *H. pylori* secreted polypeptide or a fragment thereof, wherein the polypeptide is selected from the group consisting of SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 107, SEQ ID NO: 109, SEQ ID NO: 117, SEQ ID NO: 122, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 142, SEQ ID NO: 143, SEQ ID NO: 150 SEQ ID NO: 161, SEQ ID NO: 164, SEQ ID NO: 167, SEQ ID NO: 174, SEQ ID NO: 175, SEQ ID NO: 178, SEQ ID NO: 179, SEQ ID NO: 187, SEQ ID NO: 192, and SEQ ID NO: 194.

Particularly preferred is an isolated *H. pylori* secreted polypeptide or a fragment thereof, wherein the polypeptide is encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 20, SEQ ID NO: 25, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 53 SEQ ID NO: 64, SEQ ID NO: 67, SEQ ID NO: 70, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 81, SEQ ID NO: 82, SEQ ID NO: 90, SEQ ID NO: 95, and SEQ ID NO: 97.

In another aspect, the invention features a chimeric *H. pylori* polypeptide comprising at least two *H. pylori* polypeptides or fragments thereof, wherein the polypeptides are encoded by nucleic acid sequences selected from the group consisting of SEQ ID NO:1-SEQ ID NO:97.

In another aspect, the invention features a chimeric *H. pylori* polypeptide comprising at least two *H. pylori* polypeptides or fragments thereof, wherein the polypeptides are selected from the group consisting of SEQ ID NO:98-SEQ ID NO:194.

In another aspect, the invention features a fusion protein comprising an *H. pylori* polypeptide which comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 98-SEQ ID NO: 194 operatively linked to a non-*H. pylori* polypeptide.

In another aspect, the invention features a vaccine formulation for prophylactic or therapeutic treatment of an *H. pylori* infection comprising an effective amount of at least one isolated nucleic acid of the invention.

In another aspect, the invention features a vaccine formulation for prophylactic or therapeutic treatment of an *H. pylori* infection comprising an effective amount of at least one *H. pylori* polypeptide of the invention.

Preferably, the vaccine formulation of the invention further includes a pharmaceutically acceptable carrier. In one embodiment, the pharmaceutically acceptable carrier includes an adjuvant. In another embodiment, the pharmaceutically acceptable carrier includes a delivery system, e.g., a live vector, e.g., a bacteria or a virus. In another embodiment, the pharmaceutically acceptable carrier includes both an adjuvant and a delivery system.

In another aspect, the invention features a method of treating or reducing a risk of *H. pylori* infection in a subject. The method includes administering to a subject a vaccine formulation of the invention, such that treatment or reduction of risk of *H. pylori* infection occurs.

In another aspect, the invention features a method of producing a vaccine formulation of the invention. The method includes combining at least one isolated *H. pylori* polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 98-SEQ ID NO: 194 with a pharmaceutically acceptable carrier to thereby form a vaccine formulation.

In another aspect, the invention features a method of producing a vaccine formulation of the invention. The method includes culturing a cell under condition that permit expression of an *H. pylori* polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 98-SEQ ID NO: 194; isolating the *H. pylori* polypeptide from the cell; and combining at least one isolated *H. pylori* polypeptide or a fragment thereof with a pharmaceutically acceptable carrier to thereby form a vaccine formulation.

In another aspect, the invention pertains to any individual *H. pylori* polypeptide member or nucleic acid encoding such a member from the above-identified groups of *H. pylori* polypeptides.

In another aspect, the invention features nucleic acids capable of binding mRNA of *H. pylori*. Such nucleic acid is capable of acting as antisense nucleic acid to control the translation of mRNA of *H. pylori*. A further aspect features a nucleic acid which is capable of binding specifically to an *H. pylori* nucleic acid. These nucleic acids are also referred to herein as complements and have utility as probes and as capture reagents.

In another aspect, the invention features an expression system comprising an open reading frame corresponding to *H. pylori* nucleic acid. The nucleic acid further comprises a control sequence compatible with an intended host. The expression system is useful for making polypeptides corresponding to *H. pylori* nucleic acid.

In another aspect, the invention features a cell transformed with the expression system to produce *H. pylori* polypeptides.

In another aspect, the invention features a method of generating antibodies against *H. pylori* polypeptides which are capable of binding specifically to *H. pylori* polypeptides. Such antibodies have utility as reagents for immunoassays to evaluate the abundance and distribution of *H. pylori*-specific antigens.

In another aspect, the invention features a method of generating vaccines for immunizing an individual against *H. pylori*. The vaccination method includes: immunizing a subject with at least one *H. pylori* polypeptide according to the present

invention, e.g., a surface or secreted polypeptide, or active portion thereof, and a pharmaceutically acceptable carrier. Such vaccines have therapeutic and/or prophylactic utilities.

In another aspect, the invention provides a method for generating a vaccine comprising a modified immunogenic *H. pylori* polypeptide, e.g., a surface or secreted polypeptide, or active portion thereof, and a pharmacologically acceptable carrier.

In another aspect, the invention features a method of evaluating a compound, e.g. a polypeptide, e.g., a fragment of a host cell polypeptide, for the ability to bind an *H. pylori* polypeptide. The method includes: contacting the candidate compound with an *H. pylori* polypeptide and determining if the compound binds or otherwise interacts with an *H. pylori* polypeptide. Compounds which bind *H. pylori* are candidates as activators or inhibitors of the bacterial life cycle. These assays can be performed *in vitro* or *in vivo*.

In another aspect, the invention features a method of evaluating a compound, e.g. a polypeptide, e.g., a fragment of a host cell polypeptide, for the ability to bind an *H. pylori* nucleic acid, e.g., DNA or RNA. The method includes: contacting the candidate compound with an *H. pylori* nucleic acid and determining if the compound binds or otherwise interacts with an *H. pylori* polypeptide. Compounds which bind *H. pylori* are candidates as activators or inhibitors of the bacterial life cycle. These assays can be performed *in vitro* or *in vivo*.

The invention features *H. pylori* polypeptides, preferably a substantially pure preparation of an *H. pylori* polypeptide, or a recombinant *H. pylori* polypeptide. In preferred embodiments: the polypeptide has biological activity; the polypeptide has an amino acid sequence at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% identical or homologous to an amino acid sequence of the invention contained in the Sequence Listing, preferably it has about 65% sequence identity with an amino acid sequence of the invention contained in the Sequence Listing, and most preferably it has about 92% to about 99% sequence identity with an amino acid sequence of the invention contained in the Sequence Listing; the polypeptide has an amino acid sequence essentially the same as an amino acid sequence of the invention contained in the Sequence Listing; the polypeptide is at least 5, 10, 20, 50, 100, or 150 amino acid residues in length; the polypeptide includes at least 5, preferably at least 10, more preferably at least 20, more preferably at least 50, 100, or 150 contiguous amino acid residues of the invention contained in the Sequence Listing. In yet another preferred embodiment, the amino acid sequence which differs in sequence identity by about 7% to about 8% from the *H. pylori* amino acid sequences of the invention contained in the Sequence Listing is also encompassed by the invention.

In preferred embodiments: the *H. pylori* polypeptide is encoded by a nucleic acid of the invention contained in the Sequence Listing, or by a nucleic acid having at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% homology with a nucleic acid of the invention contained in the Sequence Listing.

In a preferred embodiment, the subject *H. pylori* polypeptide differs in amino acid sequence at 1, 2, 3, 5, 10 or more residues from a sequence of the invention contained in the Sequence Listing. The differences, however, are such that the *H. pylori* polypeptide exhibits an *H. pylori* biological activity, e.g., the *H. pylori* polypeptide retains a biological activity of a naturally occurring *H. pylori* polypeptide.

In preferred embodiments, the polypeptide includes all or a fragment of an amino acid sequence of the invention contained in the Sequence Listing; fused, in reading frame, to additional amino acid residues, preferably to residues encoded by genomic DNA 5' or 3' to the genomic DNA which encodes a sequence of the invention contained in the Sequence Listing.

In yet other preferred embodiments, the *H. pylori* polypeptide is a recombinant fusion protein having a first *H. pylori* polypeptide portion and a second polypeptide portion, e.g., a second polypeptide portion having an amino acid sequence unrelated to *H. pylori*. The second polypeptide portion can be, e.g., any of glutathione-S-transferase, a DNA binding domain, or a polymerase activating domain. In preferred embodiment the fusion protein can be used in a two-hybrid assay.

Polypeptides of the invention include those which arise as a result of alternative transcription events, alternative RNA splicing events, and alternative translational and postranslational events.

The invention also encompasses an immunogenic component which includes at least one *H. pylori* polypeptide in an immunogenic preparation; the immunogenic component being capable of eliciting an immune response specific for the *H. pylori* polypeptide, e.g., a humoral response, an antibody response, or a cellular response. In preferred embodiments, the immunogenic component comprises at least one antigenic determinant from a polypeptide of the invention contained in the Sequence Listing.

In another aspect, the invention provides a substantially pure nucleic acid having a nucleotide sequence which encodes an *H. pylori* polypeptide. In preferred embodiments: the encoded polypeptide has biological activity; the encoded polypeptide has an amino acid sequence at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% homologous to an amino acid sequence of the invention contained in the Sequence Listing; the encoded polypeptide has an amino acid sequence essentially the same as an amino acid sequence of the invention contained in the Sequence Listing; the encoded polypeptide is at least 5, 10, 20, 50, 100, or 150 amino acids in length; the encoded

polypeptide comprises at least 5, preferably at least 10, more preferably at least 20, more preferably at least 50, 100, or 150 contiguous amino acids of the invention contained in the Sequence Listing.

In preferred embodiments: the nucleic acid of the invention is that contained in the Sequence Listing; the nucleic acid is at least 60%, 70%, 80%, 90%, 95%, 98%, or 99% homologous with a nucleic acid sequence of the invention contained in the Sequence Listing.

In a preferred embodiment, the encoded *H. pylori* polypeptide differs (e.g., by amino acid substitution, addition or deletion of at least one amino acid residue) in amino acid sequence at 1, 2, 3, 5, 10 or more residues, from a sequence of the invention contained in the Sequence Listing. The differences, however, are such that: the *H. pylori* encoded polypeptide exhibits a *H. pylori* biological activity, e.g., the encoded *H. pylori* enzyme retains a biological activity of a naturally occurring *H. pylori*.

In preferred embodiments, the encoded polypeptide includes all or a fragment of an amino acid sequence of the invention contained in the Sequence Listing; fused, in reading frame, to additional amino acid residues, preferably to residues encoded by genomic DNA 5' or 3' to the genomic DNA which encodes a sequence of the invention contained in the Sequence Listing.

In preferred embodiments, the subject *H. pylori* nucleic acid will include a transcriptional regulatory sequence, e.g. at least one of a transcriptional promoter or transcriptional enhancer sequence, operably linked to the *H. pylori* gene sequence, e.g., to render the *H. pylori* gene sequence suitable for expression in a recombinant host cell.

In yet a further preferred embodiment, the nucleic acid which encodes an *H. pylori* polypeptide of the invention, hybridizes under stringent conditions to a nucleic acid probe corresponding to at least 8 consecutive nucleotides of the invention contained in the Sequence Listing; more preferably to at least 12 consecutive nucleotides of the invention contained in the Sequence Listing; more preferably to at least 20 consecutive nucleotides of the invention contained in the Sequence Listing; more preferably to at least 40 consecutive nucleotides of the invention contained in the Sequence Listing.

In a preferred embodiment, the nucleic acid encodes a peptide which differs by at least one amino acid residue from the sequences of the invention contained in the Sequence Listing.

In a preferred embodiment, the nucleic acid differs by at least one nucleotide from a nucleotide sequence of the invention contained in the Sequence Listing which encodes amino acids of the invention contained in the Sequence Listing.

In another aspect, the invention encompasses: a vector including a nucleic acid which encodes an *H. pylori* polypeptide or an *H. pylori* polypeptide variant as described

herein; a host cell transfected with the vector; and a method of producing a recombinant *H. pylori* polypeptide or *H. pylori* polypeptide variant; including culturing the cell, e.g., in a cell culture medium, and isolating the *H. pylori* or *H. pylori* polypeptide variant, e.g., from the cell or from the cell culture medium.

In another aspect, the invention features, a purified recombinant nucleic acid having at least 50%, 60%, 70%, 80%, 90%, 95%, 98%, or 99% homology with a sequence of the invention contained in the Sequence Listing.

The invention also provides a probe or primer which includes a substantially purified oligonucleotide. The oligonucleotide includes a region of nucleotide sequence which hybridizes under stringent conditions to at least 8 consecutive nucleotides of sense or antisense sequence of the invention contained in the Sequence Listing, or naturally occurring mutants thereof. In preferred embodiments, the probe or primer further includes a label group attached thereto. The label group can be, e.g., a radioisotope, a fluorescent compound, an enzyme, and/or an enzyme co-factor. Preferably the oligonucleotide is at least 8 and less than 10, 20, 30, 50, 100, or 150 nucleotides in length.

The invention also provides an isolated *H. pylori* polypeptide which is encoded by a nucleic acid which hybridizes under stringent hybridization conditions to a nucleic acid contained in the Sequence Listing.

The invention further provides nucleic acids, e.g., RNA or DNA, encoding a polypeptide of the invention. This includes double stranded nucleic acids as well as coding and antisense single strands.

The *H. pylori* strain, from which genomic sequences have been sequenced, has been deposited in the American Type Culture Collection (ATCC # 55679; deposited by Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154) as strain HP-J99.

Included in the invention are: allelic variations; natural mutants; induced mutants; proteins encoded by DNA that hybridizes under high or low stringency conditions to a nucleic acid which encodes a polypeptide of the invention contained in the Sequence Listing (for definitions of high and low stringency see Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1989, 6.3.1 - 6.3.6 and 6.4.1-6.4.10, hereby incorporated by reference); and, polypeptides specifically bound by antisera to *H. pylori* polypeptides, especially by antisera to an active site or binding domain of *H. pylori* polypeptide. The invention also includes fragments, preferably biologically active fragments. These and other polypeptides are also referred to herein as *H. pylori* polypeptide analogs or variants.

Putative functions have been determined for several of the *H. pylori* polypeptides of the invention, as shown in Table 1.

Accordingly, uses of the claimed *H. pylori* polypeptides based on these identified functions, as well as other functions as described herein, are also within the scope of the invention.

In addition, the present invention encompasses *H. pylori* polypeptides characterized as shown in Table 1 below, including: *H. pylori* cell envelope proteins, *H. pylori* secreted proteins, *H. pylori* cytoplasmic proteins and *H. pylori* cellular proteins. Members of these groups were identified by BLAST homology searches and by searches for secretion signal or transmembrane protein motifs. Polypeptides related by significant homology to the polypeptides of Table 1 are also considered to be classified in the manner of the homologs shown in Table 1.

TABLE 1

	ntSeqID[PCT]	aaSeqID[PCT]
A. CELL ENVELOPE		
A.1 Flagella-associated		
hp1p13939_24322162_f3_17	63	160
A.2 Outer membrane		
A.2.1 Terminal phe residue		
02ge10116_23462_f2_43	7	104
02ge10116_804550_f2_44	8	105
02ge41622_14875000_c2_65	9	106
01cp20708_214843_c2_49	13	110
01cp20708_4960952_c1_43	14	111
06ae11016_4729625_c3_68	23	120
06ep10615_49068_c2_87	24	121
06gp71906_35158328_f3_85	27	124
06gp71906_3941642_f2_70	28	125
13ae10610_156411_c3_33	50	147
13ae10610_6522827_c3_37	51	148
hp4e53394_11798952_c2_101	61	158
06ge20501_4298568_c3_53	79	176
11ae12004_3367666_c2_41	80	177
hp7e10433_5345837_c3_13	84	181
14ce61516_2460[illegible]_f2_9	85	182
11ap20714_2077_c3_103	91	188
02cp10615_21908138_f1_4	94	191
A.2.2 No terminal phe residue		
07gp11909_26460892_f2_6	5	102
A.2.3 Phe and Tyr cluster at C-terminus		
02ge41622_34176513_c1_50	11	108

06gp71906_20486556_f2_65	26	123
hp7e10520_14728137_f1_1	36	133
02ae31010_417818_f3_29	42	139
13ae10610_26855313_f3_15	52	149
A.2.4 Via homology		
hp5p15212_13729635_c3_35	22	119
07ee11402_1046877_c3_100	29	126
14ee41924_1046877_c3_104	30	127
hp1p13939_21641016_f1_1	65	162
hp4p62853_4766691_f3_23	66	163
A.3 Inner membrane		
A.3.1 Proteins involved in transport		
06cp30603_664083_c1_94	48	145
09cp10713_36359687_c1_119	49	146
04ep41903_16667055_c1_37	17	114
04ep41903_19689182_c1_43	18	115
14ce31519_24650009_c1_17	19	116
09ce10413_26734687_f3_23	43	140
hp6p10904_6726062_f3_13	44	141
A.3.2 Other inner membrane proteins		
02ae31010_16679640_f2_21	38	135
07ee50709_16679640_f3_60	39	136
A.4 Other cell envelope proteins		
01ce61016_1056562_c3_123	1	98
09cp61003_16619192_c2_83	2	99
02ge10116_15632000_c2_114	6	103
04ae61517_12345837_f2_4	34	131
04ae61517_21744091_f3_5	35	132
hp4e13394_26750068_c3_113	60	157
hp5p15575_1053590_c1_35	69	166
hp7e10433_5345837_c2_8	83	180
B. CYTOPLASMIC PROTEINS		
B.1. Proteins involved in mRNA translation		
hp3e10946_32609412_f3_4	57	154
hp3e10946_34175837_f3_3	58	155
B.2 Proteins involved in genome replication, transcription, recombination and repair		
14ce61516_12600937_f2_11	86	183
14cp11908_25402267_c3_104	87	184
B.3 Other cytoplasmic proteins		
05ce10910_23712780_c1_4	88	185
hp7e10192_23712780_f2_5	89	186
11ap20714_34663910_f3_29	92	189
hp8e10065_4962812_f2_18	93	190
C. SECRETED PROTEINS		
01ce61016_23593955_c3_140	3	100

09cp61003_23593955_c1_79	4	101
02ge41622_20730462_f1_19	10	107
01cp20708_10628177_c2_50	12	109
05ae30220_24415693_c3_175	20	117
06gp10409_4015687_f2_11	25	122
hp2e10911_10213593_c1_73	31	128
hp2e10911_35567005_c2_88	32	129
09ze10333_1457137_f3_11	45	142
06cp30603_10744075_c3_136	46	143
12ae10622_30273255_f1_13	53	150
05ce10208_4707035_c2_17	64	161
06ep30223_176437_c2_134	67	164
hp5p15575_26016387_f2_16	70	167
hp6p12244_4881375_c3_97	77	174
06ce20610_34647187_c2_33	78	175
hp7e10433_36339535_f3_3	81	178
hp7e10433_36339535_f3_3	82	179
hp7e10420_24391078_f1_3	90	187
02ce71018_35720091_c3_4	95	192
hp6e10363_30517031_f3_3	97	194
D. OTHER CELLULAR PROTEINS		
01ae11010_26437877_c2_52	15	112
hp4p33322_5891077_c2_45	16	113
hp3p21118_54628_c3_3	21	118
02ae31010_1064125_f1_11	33	130
hp2e10911_15680337_c3_105	37	134
hp2e10911_24804577_c3_104	40	137
hp2e10911_32234750_c1_68	41	138
06cp30603_26070252_c3_140	47	144
03ae10804_235286_f3_19	54	151
09ge11604_4804692_c1_8	55	152
hp2p10610_21987687_c2_5	56	153
hp4e13394_26182793_f2_45	59	156
hp4e53394_2082126_c2_102	62	159
06ep30223_25402187_c1_112	68	165
hp6e10491_12712706_f3_12	71	168
hp6p12129_12542880_c3_29	72	169
hp6p12129_17067265_c3_29	73	170
hp6p12129_214055_f1_2	74	171
hp6p12129_214055_f3_17	75	172
hp6p12244_33492712_c3_88	76	173
hp1e13054_22360653_f2_4	96	193

[In Table 1, "nt" represents nucleotide Seq. ID number and "aa" represents amino acid Seq. ID number]

Definitions

The terms "purified polypeptide" and "isolated polypeptide" and "a substantially pure preparation of a polypeptide" are used interchangeably herein and, as used herein, mean a polypeptide that has been substantially, and preferably completely, separated from other proteins, lipids, and nucleic acids with which it naturally occurs. Preferably, the polypeptide is also separated from substances, e.g., antibodies or gel matrix, e.g., polyacrylamide, which are used to purify it. Preferably, the polypeptide constitutes at least 10, 20, 50 70, 80 or 95% dry weight of the purified preparation. Preferably, the preparation contains: sufficient polypeptide to allow protein sequencing; at least 1, 10, or 100 μg of the polypeptide; at least 1, 10, or 100 mg of the polypeptide. Furthermore, the terms "purified polypeptide" and "isolated polypeptide" and "a substantially pure preparation of a polypeptide," as used herein, refer to both a polypeptide obtained from nature or produced by recombinant DNA techniques as described herein.

For example, an "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the *H. pylori* protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of *H. pylori* protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. In one embodiment, the language "substantially free of cellular material" includes preparations of *H. pylori* protein having less than about 30% (by dry weight) of non-*H. pylori* protein (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-*H. pylori* protein, still more preferably less than about 10% of non-*H. pylori* protein, and most preferably less than about 5% non-*H. pylori* protein. When the *H. pylori* protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of *H. pylori* protein in which the protein is separated from chemical precusors or other chemicals which are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of *H. pylori* protein having less than about 30% (by dry weight) of chemical precursors or non-*H. pylori* chemicals, more preferably less than about 20% chemical precursors or non-*H. pylori* chemicals, still more preferably less

than about 10% chemical precursors or non-*H. pylori* chemicals, and most preferably less than about 5% chemical precursors or non-*H. pylori* chemicals.

A purified preparation of cells refers to, in the case of plant or animal cells, an *in vitro* preparation of cells and not an entire intact plant or animal. In the case of cultured cells or microbial cells, it consists of a preparation of at least 10% and more preferably 50% of the subject cells.

A purified or isolated or a substantially pure nucleic acid, e.g., a substantially pure DNA, (are terms used interchangeably herein) is a nucleic acid which is one or both of the following: not immediately contiguous with both of the coding sequences with which it is immediately contiguous (i.e., one at the 5' end and one at the 3' end) in the naturally-occurring genome of the organism from which the nucleic acid is derived; or which is substantially free of a nucleic acid with which it occurs in the organism from which the nucleic acid is derived. The term includes, for example, a recombinant DNA which is incorporated into a vector, e.g., into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA or a genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other DNA sequences. Substantially pure DNA also includes a recombinant DNA which is part of a hybrid gene encoding additional *H. pylori* DNA sequence.

A "contig" as used herein is a nucleic acid representing a continuous stretch of genomic sequence of an organism.

An "open reading frame", also referred to herein as ORF, is a region of nucleic acid which encodes a polypeptide. This region may represent a portion of a coding sequence or a total sequence and can be determined from a stop to stop codon or from a start to stop codon.

As used herein, a "coding sequence" is a nucleic acid which is transcribed into messenger RNA and/or translated into a polypeptide when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a translation start codon at the five prime terminus and a translation stop code at the three prime terminus. A coding sequence can include but is not limited to messenger RNA, synthetic DNA, and recombinant nucleic acid sequences.

A "complement" of a nucleic acid as used herein referes to an anti-parallel or antisense sequence that participates in Watson-Crick base-pairing with the original sequence.

A "gene product" is a protein or structural RNA which is specifically encoded by a gene.

As used herein, the term "probe" refers to a nucleic acid, peptide or other chemical entity which specifically binds to a molecule of interest. Probes are often associated with or capable of associating with a label. A label is a chemical moiety capable of detection. Typical labels comprise dyes, radioisotopes, luminescent and chemiluminescent moieties, fluorophores, enzymes, precipitating agents, amplification sequences, and the like. Similarly, a nucleic acid, peptide or other chemical entity which specifically binds to a molecule of interest and immobilizes such molecule is referred herein as a "capture ligand". Capture ligands are typically associated with or capable of associating with a support such as nitro-cellulose, glass, nylon membranes, beads, particles and the like. The specificity of hybridization is dependent on conditions such as the base pair composition of the nucleotides, and the temperature and salt concentration of the reaction. These conditions are readily discernable to one of ordinary skill in the art using routine experimentation.

Homologous refers to the sequence similarity or sequence identity between two polypeptides or between two nucleic acid molecules. When a position in both of the two compared sequences is occupied by the same base or amino acid monomer subunit, e.g., if a position in each of two DNA molecules is occupied by adenine, then the molecules are homologous at that position. The percent of homology between two sequences is a function of the number of matching or homologous positions shared by the two sequences divided by the number of positions compared x 100. For example, if 6 of 10 of the positions in two sequences are matched or homologous then the two sequences are 60% homologous. By way of example, the DNA sequences ATTGCC and TATGGC share 50% homology. Generally, a comparison is made when two sequences are aligned to give maximum homology.

Nucleic acids are hybridizable to each other when at least one strand of a nucleic acid can anneal to the other nucleic acid under defined stringency conditions. Stringency of hybridization is determined by: (a) the temperature at which hybridization and/or washing is performed; and (b) the ionic strength and polarity of the hybridization and washing solutions. Hybridization requires that the two nucleic acids contain complementary sequences; depending on the stringency of hybridization, however, mismatches may be tolerated. Typically, hybridization of two sequences at high stingency (such as, for example, in a solution of 0.5X SSC, at 65° C) requires that the sequences be essentially completely homologous. Conditions of intermediate stringency (such as, for example, 2X SSC at 65 ° C) and low stringency (such as, for example 2X SSC at 55° C), require correspondingly less overall complementarity between the hybridizing sequences. (1X SSC is 0.15 M NaCl, 0.015 M Na citrate). A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X

sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65°C.

The terms peptides, proteins, and polypeptides are used interchangeably herein.

As used herein, the term "surface protein" refers to all surface accessible proteins, e.g. inner and outer membrane proteins, proteins adhering to the cell wall, and secreted proteins.

A polypeptide has *H. pylori* biological activity if it has one, two and preferably more of the following properties: (1) if when expressed in the course of an *H. pylori* infection, it can promote, or mediate the attachment of *H. pylori* to a cell; (2) it has an enzymatic activity, structural or regulatory function characteristic of an *H. pylori* protein; (3) the gene which encodes it can rescue a lethal mutation in an *H. pylori* gene; (4) or it is immunogenic in a subject. A polypeptide has biological activity if it is an antagonist, agonist, or super-agonist of a polypeptide having one of the above-listed properties.

A biologically active fragment or analog is one having an *in vivo* or *in vitro* activity which is characteristic of the *H. pylori* polypeptides of the invention contained in the Sequence Listing, or of other naturally occurring *H. pylori* polypeptides, e.g., one or more of the biological activities described herein. Especially preferred are fragments which exist *in vivo*, e.g., fragments which arise from post transcriptional processing or which arise from translation of alternatively spliced RNA's. Fragments include those expressed in native or endogenous cells as well as those made in expression systems, e.g., in CHO cells. Because peptides such as *H. pylori* polypeptides often exhibit a range of physiological properties and because such properties may be attributable to different portions of the molecule, a useful *H. pylori* fragment or *H. pylori* analog is one which exhibits a biological activity in any biological assay for *H. pylori* activity. Most preferably the fragment or analog possesses 10%, preferably 40%, more preferably 60%, 70%, 80% or 90% or greater of the activity of *H. pylori*, in any *in vivo* or *in vitro* assay.

Analogs can differ from naturally occurring *H. pylori* polypeptides in amino acid sequence or in ways that do not involve sequence, or both. Non-sequence modifications include changes in acetylation, methylation, phosphorylation, carboxylation, or glycosylation. Preferred analogs include *H. pylori* polypeptides (or biologically active fragments thereof) whose sequences differ from the wild-type sequence by one or more conservative amino acid substitutions or by one or more non-conservative amino acid substitutions, deletions, or insertions which do not substantially diminish the biological activity of the *H. pylori* polypeptide. Conservative substitutions typically include the substitution of one amino acid for another with similar characteristics, e.g., substitutions within the following groups: valine, glycine; glycine, alanine; valine, isoleucine,

leucine; aspartic acid, glutamic acid; asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. Other conservative substitutions can be made in view of the table below.

TABLE 2

CONSERVATIVE AMINO ACID REPLACEMENTS

For Amino Acid	Code	Replace with any of
Alanine	A	D-Ala, Gly, beta-Ala, L-Cys, D-Cys
Arginine	R	D-Arg, Lys, D-Lys, homo-Arg, D-homo-Arg, Met, Ile, D-Met, D-Ile, Orn, D-Orn
Asparagine	N	D-Asn, Asp, D-Asp, Glu, D-Glu, Gln, D-Gln
Aspartic Acid	D	D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln
Cysteine	C	D-Cys, S-Me-Cys, Met, D-Met, Thr, D-Thr
Glutamine	Q	D-Gln, Asn, D-Asn, Glu, D-Glu, Asp, D-Asp
Glutamic Acid	E	D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln
Glycine	G	Ala, D-Ala, Pro, D-Pro, β-Ala, Acp
Isoleucine	I	D-Ile, Val, D-Val, Leu, D-Leu, Met, D-Met
Leucine	L	D-Leu, Val, D-Val, Leu, D-Leu, Met, D-Met
Lysine	K	D-Lys, Arg, D-Arg, homo-Arg, D-homo-Arg, Met, D-Met, Ile, D-Ile, Orn, D-Orn
Methionine	M	D-Met, S-Me-Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val
Phenylalanine	F	D-Phe, Tyr, D-Thr, L-Dopa, His, D-His, Trp, D-Trp, Trans-3,4, or 5-phenylproline, cis-3,4, or 5-phenylproline
Proline	P	D-Pro, L-I-thioazolidine-4-carboxylic acid, D-or L-1-oxazolidine-4-carboxylic acid
Serine	S	D-Ser, Thr, D-Thr, allo-Thr, Met, D-Met, Met(O), D-Met(O), L-Cys, D-Cys
Threonine	T	D-Thr, Ser, D-Ser, allo-Thr, Met, D-Met, Met(O), D-Met(O), Val, D-Val
Tyrosine	Y	D-Tyr, Phe, D-Phe, L-Dopa, His, D-His
Valine	V	D-Val, Leu, D-Leu, Ile, D-Ile, Met, D-Met

Other analogs within the invention are those with modifications which increase peptide stability; such analogs may contain, for example, one or more non-peptide bonds (which replace the peptide bonds) in the peptide sequence. Also included are: analogs that include residues other than naturally occurring L-amino acids, e.g., D-amino acids or non-naturally occurring or synthetic amino acids, e.g., β or γ amino acids; and cyclic analogs.

As used herein, the term "fragment", as applied to an *H. pylori* analog, will ordinarily be at least about 20 residues, more typically at least about 40 residues,

preferably at least about 60 residues in length. Fragments of *H. pylori* polypeptides can be generated by methods known to those skilled in the art. The ability of a candidate fragment to exhibit a biological activity of *H. pylori* polypeptide can be assessed by methods known to those skilled in the art as described herein. Also included are *H. pylori* polypeptides containing residues that are not required for biological activity of the peptide or that result from alternative mRNA splicing or alternative protein processing events.

An "immunogenic component" as used herein is a moiety, such as an *H. pylori* polypeptide, analog or fragment thereof, that is capable of eliciting a humoral and/or cellular immune response in a host animal alone or in combination with an adjuvant.

An "antigenic component" as used herein is a moiety, such as an *H. pylori* polypeptide, analog or fragment thereof, that is capable of binding to a specific antibody with sufficiently high affinity to form a detectable antigen-antibody complex.

As used herein, the term "transgene" means a nucleic acid (encoding, e.g., one or more polypeptides), which is partly or entirely heterologous, i.e., foreign, to the transgenic animal or cell into which it is introduced, or, is homologous to an endogenous gene of the transgenic animal or cell into which it is introduced, but which is designed to be inserted, or is inserted, into the cell's genome in such a way as to alter the genome of the cell into which it is inserted (e.g., it is inserted at a location which differs from that of the natural gene or its insertion results in a knockout). A transgene can include one or more transcriptional regulatory sequences and any other nucleic acid, such as introns, that may be necessary for optimal expression of the selected nucleic acid, all operably linked to the selected nucleic acid, and may include an enhancer sequence.

As used herein, the term "transgenic cell" refers to a cell containing a transgene.

As used herein, a "transgenic animal" is any animal in which one or more, and preferably essentially all, of the cells of the animal includes a transgene. The transgene can be introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by a process of transformation of competent cells or by microinjection or by infection with a recombinant virus. This molecule may be integrated within a chromosome, or it may be extrachromosomally replicating DNA.

The term "antibody" as used herein is intended to include fragments thereof which are specifically reactive with *H. pylori* polypeptides.

As used herein, the term "cell-specific promoter" means a DNA sequence that serves as a promoter, i.e., regulates expression of a selected DNA sequence operably linked to the promoter, and which effects expression of the selected DNA sequence in specific cells of a tissue. The term also covers so-called "leaky" promoters, which

regulate expression of a selected DNA primarily in one tissue, but cause expression in other tissues as well.

Misexpression, as used herein, refers to a non-wild type pattern of gene expression. It includes: expression at non-wild type levels, i.e., over or under expression; a pattern of expression that differs from wild type in terms of the time or stage at which the gene is expressed, e.g., increased or decreased expression (as compared with wild type) at a predetermined developmental period or stage; a pattern of expression that differs from wild type in terms of decreased expression (as compared with wild type) in a predetermined cell type or tissue type; a pattern of expression that differs from wild type in terms of the splicing size, amino acid sequence, post-transitional modification, or biological activity of the expressed polypeptide; a pattern of expression that differs from wild type in terms of the effect of an environmental stimulus or extracellular stimulus on expression of the gene, e.g., a pattern of increased or decreased expression (as compared with wild type) in the presence of an increase or decrease in the strength of the stimulus.

As used herein, "host cells" and other such terms denoting microorganisms or higher eukaryotic cell lines cultured as unicellular entities refers to cells which can become or have been used as recipients for a recombinant vector or other transfer DNA, and include the progeny of the original cell which has been transfected. It is understood by individuals skilled in the art that the progeny of a single parental cell may not necessarily be completely identical in genomic or total DNA compliment to the original parent, due to accident or deliberate mutation.

As used herein, the term "control sequence" refers to a nucleic acid having a base sequence which is recognized by the host organism to effect the expression of encoded sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include a promoter, ribosomal binding site, terminators, and in some cases operators; in eukaryotes, generally such control sequences include promoters, terminators and in some instances, enhancers. The term control sequence is intended to include at a minimum, all components whose presence is necessary for expression, and may also include additional components whose presence is advantageous, for example, leader sequences.

As used herein, the term "operably linked" refers to sequences joined or ligated to function in their intended manner. For example, a control sequence is operably linked to coding sequence by ligation in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequence and host cell.

The metabolism of a substance, as used herein, means any aspect of the, expression, function, action, or regulation of the substance. The metabolism of a substance includes modifications, e.g., covalent or non-covalent modifications of the substance. The metabolism of a substance includes modifications, e.g., covalent or non-covalent modification, the substance induces in other substances. The metabolism of a substance also includes changes in the distribution of the substance. The metabolism of a substance includes changes the substance induces in the distribution of other substances.

A "sample" as used herein refers to a biological sample, such as, for example, tissue or fluid isloated from an individual (including without limitation plasma, serum, cerebrospinal fluid, lymph, tears, saliva and tissue sections) or from *in vitro* cell culture constituents, as well as samples from the environment.

The practice of the invention will employ, unless otherwise indicated, conventional techniques of chemistry, molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature. *See e.g.*, Sambrook, Fritsch, and Maniatis, *Molecular Cloning; Laboratory Manual* 2nd ed. (1989); *DNA Cloning*, Volumes I and II (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds. 1984); the series, *Methods in Enzymology* (Academic Press, Inc.), particularly Vol. 154 and Vol. 155 (Wu and Grossman, eds.) and *PCR-A Practical Approach* (McPherson, Quirke, and Taylor, eds., 1991).

I. Isolation of Nucleic Acids of *H. pylori* and Uses Therefor

H. pylori Genomic Sequence

This invention provides nucleotide sequences of the genome of *H. pylori* which thus comprises a DNA sequence library of *H. pylori* genomic DNA. The detailed description that follows provides nucleotide sequences of *H. pylori*, and also describes how the sequences were obtained and how ORFs and protein-coding sequences were identified. Also described are methods of using the disclosed *H. pylori* sequences in methods including diagnostic and therapeutic applications. Furthermore, the library can be used as a database for identification and comparison of medically important sequences in this and other strains of *H. pylori*.

To determine the genomic sequence of *H. pylori*, DNA was isolated from a strain of *H. pylori* (ATCC # 55679; deposited by Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154) and mechanically sheared by nebulization to a

median size of 2 kb. Following size fractionation by gel electrophoresis, the fragments were blunt-ended, ligated to adapter oligonucleotides, and cloned into each of 20 different pMPX vectors (Rice et al., abstracts of Meeting of Genome Mapping and Sequencing, Cold Spring Harbor, NY, 5/11-5/15, 1994, p. 225) to construct a series of "shotgun" subclone libraries.

DNA sequencing was achieved using multiplex sequencing procedures essentially as disclosed in Church et al., 1988, *Science* 240:185; U.S. Patents No. 4,942,124 and 5,149,625). DNA was extracted from pooled cultures and subjected to chemical or enzymatic sequencing. Sequencing reactions were resolved by electrophoresis, and the products were transferred and covalently bound to nylon membranes. Finally, the membranes were sequentially hybridized with a series of labelled oligonucleotides complimentary to "tag" sequences present in the different shotgun cloning vectors. In this manner, a large number of sequences could be obtained from a single set of sequencing reactions. The cloning and sequencing procedures are described in more detail in the Exemplification.

Individual sequence reads obtained in this manner were assembled using the FALCON™ program (Church *et al.*, 1994, *Automated DNA Sequencing and Analysis*, J.C. Venter, ed., Academic Press) and PHRAP (P. Green, Abstracts of DOE Human Genome Program Contractor-Grantee Workshop V, Jan. 1996, p.157). The average contig length was about 3-4 kb.

A variety of approaches are used to order the contigs so as to obtain a continuous sequence representing the entire *H. pylori* genome. Synthetic oligonucleotides are designed that are complementary to sequences at the end of each contig. These oligonucleotides may be hybridized to libaries of *H. pylori* genomic DNA in, for example, lambda phage vectors or plasmid vectors to identify clones that contain sequences corresponding to the junctional regions between individual contigs. Such clones are then used to isolate template DNA and the same oligonucleotides are used as primers in polymerase chain reaction (PCR) to amplify junctional fragments, the nucleotide sequence of which is then determined.

The *H. pylori* sequences were analyzed for the presence of open reading frames (ORFs) comprising at least 180 nucleotides As a result of the analysis of ORFs based on stop-to-stop codon reads, it should be understood that these ORFs may not correspond to the ORF of a naturally-occurring *H. pylori* polypeptide. These ORFs may contain start codons which indicate the initiation of protein synthesis of a naturally-occurring *H. pylori* polypeptide. Such start codons within the ORFs provided herein can be identified by those of ordinary skill in the relevant art, and the resulting ORF and the encoded *H. pylori* polypeptide is within the scope of this invention. For example, within

the ORFs a codon such as AUG or GUG (encoding methionine or valine) which is part of the initiation signal for protein synthesis can be identified and the ORF modified to correspond to a naturally-occurring *H. pylori* polypeptide. The predicted coding regions were defined by evaluating the coding potential of such sequences with the program GENEMARK™ (Borodovsky and McIninch, 1993, *Comp. Chem.* 17:123).

Other *H. pylori* Nucleic Acids

The nucleic acids of this invention may be obtained directly from the DNA of the above referenced *H. pylori* strain by using the polymerase chain reaction (PCR). See "*PCR, A Practical Approach*" (McPherson, Quirke, and Taylor, eds., IRL Press, Oxford, UK, 1991) for details about the PCR. High fidelity PCR can be used to ensure a faithful DNA copy prior to expression. In addition, the authenticity of amplified products can be checked by conventional sequencing methods. Clones carrying the desired sequences described in this invention may also be obtained by screening the libraries by means of the PCR or by hybridization of synthetic oligonucleotide probes to filter lifts of the library colonies or plaques as known in the art (see, e.g., Sambrook et al., *Molecular Cloning, A Laboratory Manual* 2nd edition, 1989, Cold Spring Harbor Press, NY).

It is also possible to obtain nucleic acids encoding *H. pylori* polypeptides from a cDNA library in accordance with protocols herein described. A cDNA encoding an *H. pylori* polypeptide can be obtained by isolating total mRNA from an appropriate strain. Double stranded cDNAs can then be prepared from the total mRNA. Subsequently, the cDNAs can be inserted into a suitable plasmid or viral (e.g., bacteriophage) vector using any one of a number of known techniques. Genes encoding *H. pylori* polypeptides can also be cloned using established polymerase chain reaction techniques in accordance with the nucleotide sequence information provided by the invention. The nucleic acids of the invention can be DNA or RNA. Preferred nucleic acids of the invention are contained in the Sequence Listing.

The nucleic acids of the invention can also be chemically synthesized using standard techniques. Various methods of chemically synthesizing polydeoxynucleotides are known, including solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent Nos. 4,401,796 and 4,373,071, incorporated by reference herein).

Nucleic acids isolated or synthesized in accordance with features of the present invention are useful, by way of example, without limitation, as probes, primers, capture ligands, antisense genes and for developing expression systems for the synthesis of proteins and peptides corresponding to such sequences. As probes, primers, capture

ligands and antisense agents, the nucleic acid normally consists of all or part (approximately twenty or more nucleotides for specificity as well as the ability to form stable hybridization products) of the nucleic acids of the invention contained in the Sequence Listing. These uses are described in further detail below.

Probes

A nucleic acid isolated or synthesized in accordance with the sequence of the invention contained in the Sequence Listing can be used as a probe to specifically detect *H. pylori*. With the sequence information set forth in the present application, sequences of twenty or more nucleotides are identified which provide the desired inclusivity and exclusivity with respect to *H. pylori*, and extraneous nucleic acids likely to be encountered during hybridization conditions. More preferably, the sequence will comprise at least twenty to thirty nucleotides to convey stability to the hybridization product formed between the probe and the intended target molecules.

Sequences larger than 1000 nucleotides in length are difficult to synthesize but can be generated by recombinant DNA techniques. Individuals skilled in the art will readily recognize that the nucleic acids, for use as probes, can be provided with a label to facilitate detection of a hybridization product.

Nucleic acid isolated and synthesized in accordance with the sequence of the invention contained in the Sequence Listing can also be useful as probes to detect homologous regions (especially homologous genes) of other *Helicobacter* species using appropriate stringency hybridization conditions as described herein.

Capture Ligand

For use as a capture ligand, the nucleic acid selected in the manner described above with respect to probes, can be readily associated with a support. The manner in which nucleic acid is associated with supports is well known. Nucleic acid having twenty or more nucleotides in a sequence of the invention contained in the Sequence Listing have utility to separate *H. pylori* nucleic acid from the nucleic acid of each other and other organisms. Nucleic acid having twenty or more nucleotides in a sequence of the invention contained in the Sequence Listing can also have utility to separate other *Helicobacter* species from each other and from other organisms. Preferably, the sequence will comprise at least twenty nucleotides to convey stability to the hybridization product formed between the probe and the intended target molecules. Sequences larger than 1000 nucleotides in length are difficult to synthesize but can be generated by recombinant DNA techniques.

Primers

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility as primers for the amplification of *H. pylori* nucleic acid. These nucleic acids may also have utility as primers for the amplification of nucleic acids in other *Helicobacter* species. With respect to polymerase chain reaction (PCR) techniques, nucleic acid sequences of $\geq$ 10-15 nucleotides of the invention contained in the Sequence Listing have utility in conjunction with suitable enzymes and reagents to create copies of *H. pylori* nucleic acid. More preferably, the sequence will comprise twenty or more nucleotides to convey stability to the hybridization product formed between the primer and the intended target molecules. Binding conditions of primers greater than 100 nucleotides are more difficult to control to obtain specificity. High fidelity PCR can be used to ensure a faithful DNA copy prior to expression. In addition, amplified products can be checked by conventional sequencing methods.

The copies can be used in diagnostic assays to detect specific sequences, including genes from *H. pylori* and/or other *Helicobacter* species. The copies can also be incorporated into cloning and expression vectors to generate polypeptides corresponding to the nucleic acid synthesized by PCR, as is described in greater detail herein.

Antisense

Nucleic acid or nucleic acid-hybridizing derivatives isolated or synthesized in accordance with the sequences described herein have utility as antisense agents to prevent the expression of *H. pylori* genes. These sequences also have utility as antisense agents to prevent expression of genes of other *Helicobacter* species.

In one embodiment, nucleic acid or derivatives corresponding to *H. pylori* nucleic acids is loaded into a suitable carrier such as a liposome or bacteriophage for introduction into bacterial cells. For example, a nucleic acid having twenty or more nucleotides is capable of binding to bacteria nucleic acid or bacteria messenger RNA. Preferably, the antisense nucleic acid is comprised of 20 or more nucleotides to provide necessary stability of a hybridization product of non-naturally occurring nucleic acid and bacterial nucleic acid and/or bacterial messenger RNA. Nucleic acid having a sequence greater than 1000 nucleotides in length is difficult to synthesize but can be generated by recombinant DNA techniques. Methods for loading antisense nucleic acid in liposomes is known in the art as exemplified by U.S. Patent 4,241,046 issued December 23, 1980 to Papahadjopoulos et al.

II. Expression of *H. pylori* Nucleic Acids

Nucleic acid isolated or synthesized in accordance with the sequences described herein have utility to generate polypeptides. The nucleic acid of the invention exemplified in the Sequence Listing or fragments of the nucleic acid encoding active portions of *H. pylori* polypeptides can be cloned into suitable vectors or used to isolate nucleic acid. The isolated nucleic acid is combined with suitable DNA linkers and cloned into a suitable vector.

The function of a specific gene or operon can be ascertained by expression in a bacterial strain under conditions where the activity of the gene product(s) specified by the gene or operon in question can be specifically measured. Alternatively, a gene product may be produced in large quantities in an expressing strain for use as an antigen, an industrial reagent, for structural studies, etc. This expression can be accomplished in a mutant strain which lacks the activity of the gene to be tested, or in a strain that does not produce the same gene product(s). This includes, but is not limited to other *Helicobacter* strains, or other bacterial strains such as *E. coli*, *Norcardia*, *Corynebacterium*, *Campylobacter*, and *Streptomyces* species. In some cases the expression host will utilize the natural *Helicobacter* promoter whereas in others, it will be necessary to drive the gene with a promoter sequence derived from the expressing organism (e.g., an *E. coli* beta-galactosidase promoter for expression in *E. coli*).

To express a gene product using the natural *H. pylori* promoter, a procedure such as the following can be used. A restriction fragment containing the gene of interest, together with its associated natural promoter element and regulatory sequences (identified using the DNA sequence data) is cloned into an appropriate recombinant plasmid containing an origin of replication that functions in the host organism and an appropriate selectable marker. This can be accomplished by a number of procedures known to those skilled in the art. It is most preferably done by cutting the plasmid and the fragment to be cloned with the same restriction enzyme to produce compatible ends that can be ligated to join the two pieces together. The recombinant plasmid is introduced into the host organism by, for example, electroporation and cells containing the recombinant plasmid are identified by selection for the marker on the plasmid. Expression of the desired gene product is detected using an assay specific for that gene product.

In the case of a gene that requires a different promoter, the body of the gene (coding sequence) is specifically excised and cloned into an appropriate expression plasmid. This subcloning can be done by several methods, but is most easily accomplished by PCR amplification of a specific fragment and ligation into an

expression plasmid after treating the PCR product with a restriction enzyme or exonuclease to create suitable ends for cloning.

A suitable host cell for expression of a gene can be any procaryotic or eucaryotic cell. For example, an *H. pylori* polypeptide can be expressed in bacterial cells such as *E. coli*, insect cells (baculovirus), yeast, or mammalian cells such as Chinese hamster ovary cell (CHO). Other suitable host cells are known to those skilled in the art.

Expression in eucaryotic cells such as mammalian, yeast, or insect cells can lead to partial or complete glycosylation and/or formation of relevant inter- or intra-chain disulfide bonds of a recombinant peptide product. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari. et al., (1987) *Embo J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al., (1987) *Gene* 54:113-123), and pYES2 (Invitrogen Corporation, San Diego, CA). Baculovirus vectors available for expression of proteins in cultured insect cells (SF 9 cells) include the pAc series (Smith et al., (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow, V.A., and Summers, M.D., (1989) *Virology* 170:31-39). Generally, COS cells (Gluzman, Y., (1981) Cell 23:175-182) are used in conjunction with such vectors as pCDM 8 (Aruffo, A. and Seed, B., (1987) *Proc. Natl. Acad. Sci. USA* 84:8573-8577) for transient amplification/expression in mammalian cells, while CHO (dhfr⁻ Chinese Hamster Ovary) cells are used with vectors such as pMT2PC (Kaufman et al. (1987), *EMBO J.* 6:187-195) for stable amplification/expression in mammalian cells. Vector DNA can be introduced into mammalian cells via conventional techniques such as calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, or electroporation. Suitable methods for transforming host cells can be found in Sambrook et al. (Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory press (1989)), and other laboratory textbooks.

Expression in procaryotes is most often carried out in *E. coli* with either fusion or non-fusion inducible expression vectors. Fusion vectors usually add a number of NH_2 terminal amino acids to the expressed target gene. These NH_2 terminal amino acids often are referred to as a reporter group. Such reporter groups usually serve two purposes: 1) to increase the solubility of the target recombinant protein; and 2) to aid in the purification of the target recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the reporter group and the target recombinant protein to enable separation of the target recombinant protein from the reporter group subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Amrad Corp., Melbourne, Australia), pMAL (New England

Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase, maltose E binding protein, or protein A, respectively, to the target recombinant protein. A preferred reporter group is poly(His), which may be fused to the amino or carboxy terminus of the protein and which renders the recombinant fusion protein easily purifiable by metal chelate chromatography.

Inducible non-fusion expression vectors include pTrc (Amann et al., (1988) Gene 69:301-315) and pET11d (Studier et al., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 60-89). While target gene expression relies on host RNA polymerase transcription from the hybrid trp-lac fusion promoter in pTrc, expression of target genes inserted into pET11d relies on transcription from the T7 gn10-lac 0 fusion promoter mediated by coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident λ prophage harboring a T7 gn1 under the transcriptional control of the lacUV 5 promoter.

For example, a host cell transfected with a nucleic acid vector directing expression of a nucleotide sequence encoding an *H. pylori* polypeptide can be cultured under appropriate conditions to allow expression of the polypeptide to occur. The polypeptide may be secreted and isolated from a mixture of cells and medium containing the peptide. Alternatively, the polypeptide may be retained cytoplasmically and the cells harvested, lysed and the protein isolated. A cell culture includes host cells, media and other byproducts. Suitable media for cell culture are well known in the art. Polypeptides of the invention can be isolated from cell culture medium, host cells, or both using techniques known in the art for purifying proteins including ion-exchange chromatography, gel filtration chromatography, ultrafiltration, electrophoresis, and immunoaffinity purification with antibodies specific for such polypeptides. Additionally, in many situations, polypeptides can be produced by chemical cleavage of a native protein (e.g., tryptic digestion) and the cleavage products can then be purified by standard techniques.

In the case of membrane bound proteins, these can be isolated from a host cell by contacting a membrane-associated protein fraction with a detergent forming a solubilized complex, where the membrane-associated protein is no longer entirely embedded in the membrane fraction and is solubilized at least to an extent which allows it to be chromatographically isolated from the membrane fraction. Several different criteria are used for choosing a detergent suitable for solubilizing these complexes. For example, one property considered is the ability of the detergent to solubilize the *H. pylori* protein within the membrane fraction at minimal denaturation of the membrane-associated protein allowing for the activity or functionality of the membrane-associated

protein to return upon reconstitution of the protein. Another property considered when selecting the detergent is the critical micelle concentration (CMC) of the detergent in that the detergent of choice preferably has a high CMC value allowing for ease of removal after reconstitution. A third property considered when selecting a detergent is the hydrophobicity of the detergent. Typically, membrane-associated proteins are very hydrophobic and therefore detergents which are also hydrophobic, e.g., the triton series, would be useful for solubilizing the hydrophobic proteins. Another property important to a detergent can be the capability of the detergent to remove the *H. pylori* protein with minimal protein-protein interaction facilitating further purification. A fifth property of the detergent which should be considered is the charge of the detergent. For example, if it is desired to use ion exchange resins in the purification process then preferably detergent should be an uncharged detergent. Chromatographic techniques which can be used in the final purification step are known in the art and include hydrophobic interaction, lectin affinity, ion exchange, dye affinity and immunoaffinity.

One strategy to maximize recombinant *H. pylori* peptide expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, S., Gene Expression Technology: Methods in Enzymology 185, Academic Press, San Diego, California (1990) 119-128). Another strategy would be to alter the nucleic acid encoding an *H. pylori* peptide to be inserted into an expression vector so that the individual codons for each amino acid would be those preferentially utilized in highly expressed *E. coli* proteins (Wada et al., (1992) *Nuc. Acids Res.* 20:2111-2118). Such alteration of nucleic acids of the invention can be carried out by standard DNA synthesis techniques.

The nucleic acids of the invention can also be chemically synthesized using standard techniques. Various methods of chemically synthesizing polydeoxynucleotides are known, including solid-phase synthesis which, like peptide synthesis, has been fully automated in commercially available DNA synthesizers (See, e.g., Itakura et al. U.S. Patent No. 4,598,049; Caruthers et al. U.S. Patent No. 4,458,066; and Itakura U.S. Patent Nos. 4,401,796 and 4,373,071, incorporated by reference herein).

III. *H. pylori* Polypeptides

This invention encompasses isolated *H. pylori* polypeptides encoded by the disclosed *H. pylori* genomic sequences, including the polypeptides of the invention contained in the Sequence Listing. Polypeptides of the invention are preferably at least 5 amino acid residues in length. Using the DNA sequence information provided herein, the amino acid sequences of the polypeptides encompassed by the invention can be deduced using methods well-known in the art. It will be understood that the sequence of

an entire nucleic acid encoding an *H. pylori* polypeptide can be isolated and identified based on an ORF that encodes only a fragment of the cognate protein-coding region. This can be acheived, for example, by using the isolated nucleic acid encoding the ORF, or fragments thereof, to prime a polymerase chain reaction with genomic *H. pylori* DNA as template; this is followed by sequencing the amplified product.

The polypeptides of the invention can be isolated from wild-type or mutant *H. pylori* cells or from heterologous organisms or cells (including, but not limited to, bacteria, yeast, insect, plant and mammalian cells) into which an *H. pylori* nucleic acid has been introduced and expressed. In addition, the polypeptides can be part of recombinant fusion proteins.

H. pylori polypeptides of the invention can be chemically synthesized using commercially automated procedures such as those referenced herein.

H. pylori polypeptides of the invention are also intended to include chimeric proteins and truncated proteins as decribed herein.

Chimeric *H. pylori* proteins

H. pylori chimeric polypeptides comprise one or more *H. pylori* polypeptides fused together. These combined sequences can be made by combining two or more genes, or two or more polypeptide encoding sequences, or at least one gene and at least one polypeptide encoding sequence in tandem, and the subsequent expression of the encoded proteins by conventional molecular biological techniques. The combined nucleotide sequences may be composed of a combination of either full length *H. pylori* nucleotide sequences or fragments of such sequences, e.g., fragments which contain immunologically relevant portions of the encoded *H. pylori* protein. These chimeric *H. pylori* proteins then contain the combined or synergistic vaccine potential of each individual *H. pylori* protein sequence and can be used in vaccine formulations of the invention.

Truncated gene expression and protein production

H. pylori proteins encoded by a given nucleotide sequence can also be used in a biologically active truncated form. Such truncation can be produced, for example, by the elimination of either 5' and/or 3' regions of the encoding nucleotide sequence. These truncations can affect recombinant expression of the encoded protein and/or subsequent purification of the protein. For example, truncation of a nucleotide sequence encoding a predicted export sequence of a specific protein may alter expression of the protein. Alternatively, C-terminal truncation of an *H. pylori* polypeptide by elimination of the 3' end of the nucleic acid coding region may also improve protein expression and

subsequent purification and use, as is outlined in Example VIII below. Deletion of nucleic acid regions encoding internal *H. pylori* protein regions can also result in improved protein expression, purification and/or efficacy as a vaccine candidate.

IV. Identification of Nucleic Acids Encoding Vaccine Components and Targets for Agents Effective Against *H. pylori*

The disclosed *H. pylori* genome sequence includes segments that direct the synthesis of ribonucleic acids and polypeptides, as well as origins of replication, promoters, other types of regulatory sequences, and intergenic nucleic acids. The invention encompasses nucleic acids encoding immunogenic components of vaccines and targets for agents effective against *H. pylori*. Identification of said immunogenic components involved in the determination of the function of the disclosed sequences can be achieved using a variety of approaches. Non-limiting examples of these approaches are described briefly below.

Homology to known sequences: Computer-assisted comparison of the disclosed *H. pylori* sequences with previously reported sequences present in publicly available databases is useful for identifying functional *H. pylori* nucleic acid and polypeptide sequences. It will be understood that protein-coding sequences, for example, may be compared as a whole, and that a high degree of sequence homology between two proteins (such as, for example, >80-90%) at the amino acid level indicates that the two proteins also possess some degree of functional homology, such as, for example, among enzymes involved in metabolism, DNA synthesis, or cell wall synthesis, and proteins involved in transport, cell division, etc. In addition, many structural features of particular protein classes have been identified and correlate with specific consensus sequences, such as, for example, binding domains for nucleotides, DNA, metal ions, and other small molecules; sites for covalent modifications such as phosphorylation, acylation, and the like; sites of protein:protein interactions, etc. These consensus sequences may be quite short and thus may represent only a fraction of the entire protein-coding sequence. Identification of such a feature in an *H. pylori* sequence is therefore useful in determining the function of the encoded protein and identifying useful targets of antibacterial drugs.

Of particular relevance to the present invention are structural features that are common to secretory, transmembrane, and surface proteins, including secretion signal peptides and hydrophobic transmembrane domains. *H. pylori* proteins identified as containing putative signal sequences and/or transmembrane domains are useful as immunogenic components of vaccines.

Identification of essential genes: Nucleic acids that encode proteins essential for growth or viability of *H. pylori* are preferred drug targets. *H. pylori* genes can be tested for their biological relevance to the organism by examining the effect of deleting and/or disrupting the genes, i.e., by so-called gene "knockout", using techniques known to those skilled in the relevant art. In this manner, essential genes may be identified.

Strain-specific sequences: Because of the evolutionary relationship between different *H. pylori* strains, it is believed that the presently disclosed *H. pylori* sequences are useful for identifying, and/or discriminating between, previously known and new *H. pylori* strains. It is believed that other *H. pylori* strains will exhibit at least 70% sequence homology with the presently disclosed sequence. Systematic and routine analyses of DNA sequences derived from samples containing *H. pylori* strains, and comparison with the present sequence allows for the identification of sequences that can be used to discriminate between strains, as well as those that are common to all *H. pylori* strains. In one embodiment, the invention provides nucleic acids, including probes, and peptide and polypeptide sequences that discriminate between different strains of *H. pylori*. Strain-specific components can also be identified functionally by their ability to elicit or react with antibodies that selectively recognize one or more *H. pylori* strains.

In another embodiment, the invention provides nucleic acids, including probes, and peptide and polypeptide sequences that are common to all *H. pylori* strains but are *not* found in other bacterial species.

Specific Example: Determination Of Candidate Protein Antigens For Antibody And Vaccine Development

The selection of candidate protein antigens for vaccine development can be derived from the nucleic acids encoding *H. pylori* polypeptides. First, the ORF's can be analyzed for homology to other known exported or membrane proteins and analyzed using the discriminant analysis described by Klein, et al. (Klein, P., Kanehsia, M., and DeLisi, C. (1985) *Biochimica et Biophysica Acta* 815, 468-476) for predicting exported and membrane proteins.

Homology searches can be performed using the BLAST algorithm contained in the Wisconsin Sequence Analysis Package (Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711) to compare each predicted ORF amino acid sequence with all sequences found in the current GenBank, SWISS-PROT and PIR databases. BLAST searches for local alignments between the ORF and the databank sequences and reports a probability score which indicates the probability of finding this sequence by chance in the database. ORF's with significant homology (e.g. probabilities lower than $1x10^{-6}$ that the homology is only due to random chance) to

membrane or exported proteins represent protein antigens for vaccine development. Possible functions can be provided to *H. pylori* genes based on sequence homology to genes cloned in other organisms.

Discriminant analysis (Klein, et al. supra) can be used to examine the ORF amino acid sequences. This algorithm uses the intrinsic information contained in the ORF amino acid sequence and compares it to information derived from the properties of known membrane and exported proteins. This comparison predicts which proteins will be exported, membrane associated or cytoplasmic. ORF amino acid sequences identified as exported or membrane associated by this algorithm are likely protein antigens for vaccine development.

Surface exposed outer membrane proteins are likely to represent the best antigens to provide a protective immune response against *H. pylori*. Among the algorithms that can be used to aid in prediction of these outer membrane proteins include the presence of an amphipathic beta-sheet region at their C-terminus. This region which has been detected in a large number of outer membrane proteins in Gram negative bacteria is often characterized by hydrophobic residues (Phe or Tyr) approximately at positions 1, 3, 5, 7 and 9 from the C-terminus (e.g., see Figure 1, block F). Importantly, these sequences have not been detected at the C-termini of periplasmic proteins, thus allowing preliminary distinction between these classes of proteins based on primary sequence data. This phenomenon has been reported previously by Struyve et al. (*J. Mol. Biol.* 218:141-148, 1991).

Also illustrated in Figure 1 are additional amino acid sequence motifs found in many outer membrane proteins of *H. pylori*. The amino acid sequence alignment in Figure 1 depicts portions of the sequence of five *H. pylori* proteins (depicted in the single letter amino acid code) labeled with their amino acid Sequence ID Numbers and shown N-terminal to C-terminal, left to right. Six distinct blocks (labeled A through F) of similar amino acid residues are found including the distinctive hydrophobic residues (Phe or Tyr; F or Y according to the single letter code for amino acid residues) frequently found at positions near the C-terminus of outer membrane proteins. The presence of several shared motifs clearly establishes the similarity between members of this group of proteins.

In addition, outer membrane proteins isolated from *H. pylori* frequently share a motif near the mature N-terminus (i.e., after processing to remove the secretion signal) as illustrated in the blocked amino acid residues in Figure 2. Figure 2 depicts the N-terminal portion of three *H. pylori* proteins (designated by their amino acid Sequence ID Numbers and shown N-terminal to C-terminal, left to right).

One skilled in the art would know that these shared sequence motifs are highly significant and establish a similarity among this group of proteins.

Infrequently it is not possible to distinguish between multiple possible nucleotides at a given position in the nucleic acid sequence. In those cases the ambiguities are denoted by an extended alphabet as follows:

These are the official IUPAC-IUB single-letter base codes

Code	Base Description	
G	Guanine	
A	Adenine	
T	Thymine	
C	Cytosine	
R	Purine	(A or G)
Y	Pyrimidine	(C or T or U)
M	Amino	(A or C)
K	Ketone	(G or T)
S	Strong interaction	(C or G)
W	Weak interaction	(A or T)
H	Not-G	(A or C or T)
B	Not-A	(C or G or T)
V	Not-T (not-U)	(A or C or G)
D	Not-C	(A or G or T)
N	Any	(A or C or G or T)

The amino acid translations of this invention account for the ambiguity in the nucleic acid sequence by translating the ambiguous codon as the letter "X". In all cases, the permissible amino acid residues at a position are clear from an examination of the nucleic acid sequence based on the standard genetic code.

V. Production of Fragments and Analogs of *H. pylori* Nucleic Acids and Polypeptides

Based on the discovery of the *H. pylori* gene products of the invention provided in the Sequence Lsiting, one skilled in the art can alter the disclosed structure (of *H. pylori* genes), e.g., by producing fragments or analogs, and test the newly produced structures for activity. Examples of techniques known to those skilled in the relevant art which allow the production and testing of fragments and analogs are discussed below. These, or analogous methods can be used to make and screen libraries of polypeptides,

e.g., libraries of random peptides or libraries of fragments or analogs of cellular proteins for the ability to bind *H. pylori* polypeptides. Such screens are useful for the identification of inhibitors of *H. pylori*.

Generation of Fragments

Fragments of a protein can be produced in several ways, e.g., recombinantly, by proteolytic digestion, or by chemical synthesis. Internal or terminal fragments of a polypeptide can be generated by removing one or more nucleotides from one end (for a terminal fragment) or both ends (for an internal fragment) of a nucleic acid which encodes the polypeptide. Expression of the mutagenized DNA produces polypeptide fragments. Digestion with "end-nibbling" endonucleases can thus generate DNA's which encode an array of fragments. DNA's which encode fragments of a protein can also be generated by random shearing, restriction digestion or a combination of the above-discussed methods.

Fragments can also be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry. For example, peptides of the present invention may be arbitrarily divided into fragments of desired length with no overlap of the fragments, or divided into overlapping fragments of a desired length.

Alteration of Nucleic Acids and Polypeptides: Random Methods

Amino acid sequence variants of a protein can be prepared by random mutagenesis of DNA which encodes a protein or a particular domain or region of a protein. Useful methods include PCR mutagenesis and saturation mutagenesis. A library of random amino acid sequence variants can also be generated by the synthesis of a set of degenerate oligonucleotide sequences. (Methods for screening proteins in a library of variants are elsewhere herein).

(A) PCR Mutagenesis

In PCR mutagenesis, reduced Taq polymerase fidelity is used to introduce random mutations into a cloned fragment of DNA (Leung et al., 1989, *Technique* 1:11-15). The DNA region to be mutagenized is amplified using the polymerase chain reaction (PCR) under conditions that reduce the fidelity of DNA synthesis by Taq DNA polymerase, e.g., by using a dGTP/dATP ratio of five and adding Mn^{2+} to the PCR reaction. The pool of amplified DNA fragments are inserted into appropriate cloning vectors to provide random mutant libraries.

(B) Saturation Mutagenesis

Saturation mutagenesis allows for the rapid introduction of a large number of single base substitutions into cloned DNA fragments (Mayers et al., 1985, *Science* 229:242). This technique includes generation of mutations, e.g., by chemical treatment or irradiation of single-stranded DNA *in vitro,* and synthesis of a complimentary DNA strand. The mutation frequency can be modulated by modulating the severity of the treatment, and essentially all possible base substitutions can be obtained. Because this procedure does not involve a genetic selection for mutant fragments both neutral substitutions, as well as those that alter function, are obtained. The distribution of point mutations is not biased toward conserved sequence elements.

(C) Degenerate Oligonucleotides

A library of homologs can also be generated from a set of degenerate oligonucleotide sequences. Chemical synthesis of a degenerate sequences can be carried out in an automatic DNA synthesizer, and the synthetic genes then ligated into an appropriate expression vector. The synthesis of degenerate oligonucleotides is known in the art (see for example, Narang, SA (1983) *Tetrahedron* 39:3; Itakura et al. (1981) *Recombinant DNA, Proc 3rd Cleveland Sympos. Macromolecules*, ed. AG Walton, Amsterdam: Elsevier pp273-289; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477. Such techniques have been employed in the directed evolution of other proteins (see, for example, Scott et al. (1990) *Science* 249:386-390; Roberts et al. (1992) *PNAS* 89:2429-2433; Devlin et al. (1990) *Science* 249: 404-406; Cwirla et al. (1990) *PNAS* 87: 6378-6382; as well as U.S. Patents Nos. 5,223,409, 5,198,346, and 5,096,815).

Alteration of Nucleic Acids and Polypeptides: Methods for Directed Mutagenesis

Non-random or directed, mutagenesis techniques can be used to provide specific sequences or mutations in specific regions. These techniques can be used to create variants which include, e.g., deletions, insertions, or substitutions, of residues of the known amino acid sequence of a protein. The sites for mutation can be modified individually or in series, e.g., by (1) substituting first with conserved amino acids and then with more radical choices depending upon results achieved, (2) deleting the target residue, or (3) inserting residues of the same or a different class adjacent to the located site, or combinations of options 1-3.

(A) Alanine Scanning Mutagenesis

Alanine scanning mutagenesis is a useful method for identification of certain residues or regions of the desired protein that are preferred locations or domains for mutagenesis, Cunningham and Wells (*Science* 244:1081-1085, 1989). In alanine scanning, a residue or group of target residues are identified (e.g., charged residues such as Arg, Asp, His, Lys, and Glu) and replaced by a neutral or negatively charged amino acid (most preferably alanine or polyalanine). Replacement of an amino acid can affect the interaction of the amino acids with the surrounding aqueous environment in or outside the cell. Those domains demonstrating functional sensitivity to the substitutions are then refined by introducing further or other variants at or for the sites of substitution. Thus, while the site for introducing an amino acid sequence variation is predetermined, the nature of the mutation per se need not be predetermined. For example, to optimize the performance of a mutation at a given site, alanine scanning or random mutagenesis may be conducted at the target codon or region and the expressed desired protein subunit variants are screened for the optimal combination of desired activity.

(B) Oligonucleotide-Mediated Mutagenesis

Oligonucleotide-mediated mutagenesis is a useful method for preparing substitution, deletion, and insertion variants of DNA, see, e.g., Adelman et al., (*DNA* 2:183, 1983). Briefly, the desired DNA is altered by hybridizing an oligonucleotide encoding a mutation to a DNA template, where the template is the single-stranded form of a plasmid or bacteriophage containing the unaltered or native DNA sequence of the desired protein. After hybridization, a DNA polymerase is used to synthesize an entire second complementary strand of the template that will thus incorporate the oligonucleotide primer, and will code for the selected alteration in the desired protein DNA. Generally, oligonucleotides of at least 25 nucleotides in length are used. An optimal oligonucleotide will have 12 to 15 nucleotides that are completely complementary to the template on either side of the nucleotide(s) coding for the mutation. This ensures that the oligonucleotide will hybridize properly to the single-stranded DNA template molecule. The oligonucleotides are readily synthesized using techniques known in the art such as that described by Crea et al. (*Proc. Natl. Acad. Sci.* USA, 75:.5765[1978]).

(C) Cassette Mutagenesis

Another method for preparing variants, cassette mutagenesis, is based on the technique described by Wells et al. (*Gene*, 34:315[1985]). The starting material is a plasmid (or other vector) which includes the protein subunit DNA to be mutated. The

codon(s) in the protein subunit DNA to be mutated are identified. There must be a unique restriction endonuclease site on each side of the identified mutation site(s). If no such restriction sites exist, they may be generated using the above-described oligonucleotide-mediated mutagenesis method to introduce them at appropriate locations in the desired protein subunit DNA. After the restriction sites have been introduced into the plasmid, the plasmid is cut at these sites to linearize it. A double-stranded oligonucleotide encoding the sequence of the DNA between the restriction sites but containing the desired mutation(s) is synthesized using standard procedures. The two strands are synthesized separately and then hybridized together using standard techniques. This double-stranded oligonucleotide is referred to as the cassette. This cassette is designed to have 3' and 5' ends that are comparable with the ends of the linearized plasmid, such that it can be directly ligated to the plasmid. This plasmid now contains the mutated desired protein subunit DNA sequence.

(D) Combinatorial Mutagenesis

Combinatorial mutagenesis can also be used to generate mutants (Ladner et al., WO 88/06630). In this method, the amino acid sequences for a group of homologs or other related proteins are aligned, preferably to promote the highest homology possible. All of the amino acids which appear at a given position of the aligned sequences can be selected to create a degenerate set of combinatorial sequences. The variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level, and is encoded by a variegated gene library. For example, a mixture of synthetic oligonucleotides can be enzymatically ligated into gene sequences such that the degenerate set of potential sequences are expressible as individual peptides, or alternatively, as a set of larger fusion proteins containing the set of degenerate sequences.

Other Modifications of *H. pylori* Nucleic Acids and Polypeptides

It is possible to modify the structure of an *H. pylori* polypeptide for such purposes as increasing solubility, enhancing stability (e.g., shelf life *ex vivo* and resistance to proteolytic degradation *in vivo*). A modified *H. pylori* protein or peptide can be produced in which the amino acid sequence has been altered, such as by amino acid substitution, deletion, or addition as described herein.

An *H. pylori* peptide can also be modified by substitution of cysteine residues preferably with alanine, serine, threonine, leucine or glutamic acid residues to minimize dimerization via disulfide linkages. In addition, amino acid side chains of fragments of

the protein of the invention can be chemically modified. Another modification is cyclization of the peptide.

In order to enhance stability and/or reactivity, an *H. pylori* polypeptide can be modified to incorporate one or more polymorphisms in the amino acid sequence of the protein resulting from any natural allelic variation. Additionally, D-amino acids, non-natural amino acids, or non-amino acid analogs can be substituted or added to produce a modified protein within the scope of this invention. Furthermore, an *H. pylori* polypeptide can be modified using polyethylene glycol (PEG) according to the method of A. Sehon and co-workers (Wie et al., supra) to produce a protein conjugated with PEG. In addition, PEG can be added during chemical synthesis of the protein. Other modifications of *H. pylori* proteins include reduction/alkylation (Tarr, *Methods of Protein Microcharacterization*, J. E. Silver ed., Humana Press, Clifton NJ 155-194 (1986)); acylation (Tarr, supra); chemical coupling to an appropriate carrier (Mishell and Shiigi, eds, *Selected Methods in Cellular Immunology*, WH Freeman, San Francisco, CA (1980), U.S. Patent 4,939,239; or mild formalin treatment (Marsh, (1971) *Int. Arch. of Allergy and Appl. Immunol.*, 41: 199 - 215).

To facilitate purification and potentially increase solubility of an *H. pylori* protein or peptide, it is possible to add an amino acid fusion moiety to the peptide backbone. For example, hexa-histidine can be added to the protein for purification by immobilized metal ion affinity chromatography (Hochuli, E. et al., (1988) *Bio/Technology*, 6: 1321 - 1325). In addition, to facilitate isolation of peptides free of irrelevant sequences, specific endoprotease cleavage sites can be introduced between the sequences of the fusion moiety and the peptide.

To potentially aid proper antigen processing of epitopes within an *H. pylori* polypeptide, canonical protease sensitive sites can be engineered between regions, each comprising at least one epitope via recombinant or synthetic methods. For example, charged amino acid pairs, such as KK or RR, can be introduced between regions within a protein or fragment during recombinant construction thereof. The resulting peptide can be rendered sensitive to cleavage by cathepsin and/or other trypsin-like enzymes which would generate portions of the protein containing one or more epitopes. In addition, such charged amino acid residues can result in an increase in the solubility of the peptide.

Primary Methods for Screening Polypeptides and Analogs

Various techniques are known in the art for screening generated mutant gene products. Techniques for screening large gene libraries often include cloning the gene library into replicable expression vectors, transforming appropriate cells with the

resulting library of vectors, and expressing the genes under conditions in which detection of a desired activity, e.g., in this case, binding to *H. pylori* polypeptide or an interacting protein, facilitates relatively easy isolation of the vector encoding the gene whose product was detected. Each of the techniques described below is amenable to high through-put analysis for screening large numbers of sequences created, e.g., by random mutagenesis techniques.

(A) Two Hybrid Systems

Two hybrid assays such as the system described above (as with the other screening methods described herein), can be used to identify polypeptides, e.g., fragments or analogs of a naturally-occurring *H. pylori* polypeptide, e.g., of cellular proteins, or of randomly generated polypeptides which bind to an *H. pylori* protein. (The *H. pylori* domain is used as the bait protein and the library of variants are expressed as fish fusion proteins.) In an analogous fashion, a two hybrid assay (as with the other screening methods described herein), can be used to find polypeptides which bind a *H. pylori* polypeptide.

(B) Display Libraries

In one approach to screening assays, the candidate peptides are displayed on the surface of a cell or viral particle, and the ability of particular cells or viral particles to bind an appropriate receptor protein via the displayed product is detected in a "panning assay". For example, the gene library can be cloned into the gene for a surface membrane protein of a bacterial cell, and the resulting fusion protein detected by panning (Ladner et al., WO 88/06630; Fuchs et al. (1991) *Bio/Technology* 9:1370-1371; and Goward et al. (1992) *TIBS* 18:136-140). In a similar fashion, a detectably labeled ligand can be used to score for potentially functional peptide homologs. Fluorescently labeled ligands, e.g., receptors, can be used to detect homologs which retain ligand-binding activity. The use of fluorescently labeled ligands, allows cells to be visually inspected and separated under a fluorescence microscope, or, where the morphology of the cell permits, to be separated by a fluorescence-activated cell sorter.

A gene library can be expressed as a fusion protein on the surface of a viral particle. For instance, in the filamentous phage system, foreign peptide sequences can be expressed on the surface of infectious phage, thereby conferring two significant benefits. First, since these phage can be applied to affinity matrices at concentrations well over 10^{13} phage per milliliter, a large number of phage can be screened at one time. Second, since each infectious phage displays a gene product on its surface, if a particular phage is recovered from an affinity matrix in low yield, the phage can be amplified by

another round of infection. The group of almost identical *E. coli* filamentous phages M13, fd., and f1 are most often used in phage display libraries. Either of the phage gIII or gVIII coat proteins can be used to generate fusion proteins without disrupting the ultimate packaging of the viral particle. Foreign epitopes can be expressed at the NH_2-terminal end of pIII and phage bearing such epitopes recovered from a large excess of phage lacking this epitope (Ladner et al. PCT publication WO 90/02909; Garrard et al., PCT publication WO 92/09690; Marks et al. (1992) *J. Biol. Chem.* 267:16007-16010; Griffiths et al. (1993) *EMBO J* 12:725-734; Clackson et al. (1991) *Nature* 352:624-628; and Barbas et al. (1992) *PNAS* 89:4457-4461).

A common approach uses the maltose receptor of *E. coli* (the outer membrane protein, LamB) as a peptide fusion partner (Charbit et al. (1986) *EMBO* 5, 3029-3037). Oligonucleotides have been inserted into plasmids encoding the LamB gene to produce peptides fused into one of the extracellular loops of the protein. These peptides are available for binding to ligands, e.g., to antibodies, and can elicit an immune response when the cells are administered to animals. Other cell surface proteins, e.g., OmpA (Schorr et al. (1991) *Vaccines 91*, pp. 387-392), PhoE (Agterberg, et al. (1990) *Gene* 88, 37-45), and PAL (Fuchs et al. (1991) *Bio/Tech* 9, 1369-1372), as well as large bacterial surface structures have served as vehicles for peptide display. Peptides can be fused to pilin, a protein which polymerizes to form the pilus-a conduit for interbacterial exchange of genetic information (Thiry et al. (1989) *Appl. Environ. Microbiol.* 55, 984-993). Because of its role in interacting with other cells, the pilus provides a useful support for the presentation of peptides to the extracellular environment. Another large surface structure used for peptide display is the bacterial motive organ, the flagellum. Fusion of peptides to the subunit protein flagellin offers a dense array of many peptide copies on the host cells (Kuwajima et al. (1988) *Bio/Tech.* 6, 1080-1083). Surface proteins of other bacterial species have also served as peptide fusion partners. Examples include the *Staphylococcus* protein A and the outer membrane IgA protease of *Neisseria* (Hansson et al. (1992) *J. Bacteriol.* 174, 4239-4245 and Klauser et al. (1990) *EMBO J.* 9, 1991-1999).

In the filamentous phage systems and the LamB system described above, the physical link between the peptide and its encoding DNA occurs by the containment of the DNA within a particle (cell or phage) that carries the peptide on its surface. Capturing the peptide captures the particle and the DNA within. An alternative scheme uses the DNA-binding protein LacI to form a link between peptide and DNA (Cull *et al.* (1992) *PNAS USA* 89:1865-1869). This system uses a plasmid containing the LacI gene with an oligonucleotide cloning site at its 3'-end. Under the controlled induction by arabinose, a LacI-peptide fusion protein is produced. This fusion retains the natural

ability of LacI to bind to a short DNA sequence known as LacO operator (LacO). By installing two copies of LacO on the expression plasmid, the LacI-peptide fusion binds tightly to the plasmid that encoded it. Because the plasmids in each cell contain only a single oligonucleotide sequence and each cell expresses only a single peptide sequence, the peptides become specifically and stably associated with the DNA sequence that directed its synthesis. The cells of the library are gently lysed and the peptide-DNA complexes are exposed to a matrix of immobilized receptor to recover the complexes containing active peptides. The associated plasmid DNA is then reintroduced into cells for amplification and DNA sequencing to determine the identity of the peptide ligands. As a demonstration of the practical utility of the method, a large random library of dodecapeptides was made and selected on a monoclonal antibody raised against the opioid peptide dynorphin B. A cohort of peptides was recovered, all related by a consensus sequence corresponding to a six-residue portion of dynorphin B. (Cull et al. (1992) *Proc. Natl. Acad. Sci. U.S.A.* 89-1869)

This scheme, sometimes referred to as peptides-on-plasmids, differs in two important ways from the phage display methods. First, the peptides are attached to the C-terminus of the fusion protein, resulting in the display of the library members as peptides having free carboxy termini. Both of the filamentous phage coat proteins, pIII and pVIII, are anchored to the phage through their C-termini, and the guest peptides are placed into the outward-extending N-terminal domains. In some designs, the phage-displayed peptides are presented right at the amino terminus of the fusion protein. (Cwirla, et al. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 6378-6382) A second difference is the set of biological biases affecting the population of peptides actually present in the libraries. The LacI fusion molecules are confined to the cytoplasm of the host cells. The phage coat fusions are exposed briefly to the cytoplasm during translation but are rapidly secreted through the inner membrane into the periplasmic compartment, remaining anchored in the membrane by their C-terminal hydrophobic domains, with the N-termini, containing the peptides, protruding into the periplasm while awaiting assembly into phage particles. The peptides in the LacI and phage libraries may differ significantly as a result of their exposure to different proteolytic activities. The phage coat proteins require transport across the inner membrane and signal peptidase processing as a prelude to incorporation into phage. Certain peptides exert a deleterious effect on these processes and are underrepresented in the libraries (Gallop et al. (1994) *J. Med. Chem.* 37(9):1233-1251). These particular biases are not a factor in the LacI display system.

The number of small peptides available in recombinant random libraries is enormous. Libraries of 10^7-10^9 independent clones are routinely prepared. Libraries as

large as 10^{11} recombinants have been created, but this size approaches the practical limit for clone libraries. This limitation in library size occurs at the step of transforming the DNA containing randomized segments into the host bacterial cells. To circumvent this limitation, an *in vitro* system based on the display of nascent peptides in polysome complexes has recently been developed. This display library method has the potential of producing libraries 3-6 orders of magnitude larger than the currently available phage/phagemid or plasmid libraries. Furthermore, the construction of the libraries, expression of the peptides, and screening, is done in an entirely cell-free format.

In one application of this method (Gallop et al. (1994) *J. Med. Chem.* 37(9):1233-1251), a molecular DNA library encoding 10^{12} decapeptides was constructed and the library expressed in an *E. coli* S30 *in vitro* coupled transcription/translation system. Conditions were chosen to stall the ribosomes on the mRNA, causing the accumulation of a substantial proportion of the RNA in polysomes and yielding complexes containing nascent peptides still linked to their encoding RNA. The polysomes are sufficiently robust to be affinity purified on immobilized receptors in much the same way as the more conventional recombinant peptide display libraries are screened. RNA from the bound complexes is recovered, converted to cDNA, and amplified by PCR to produce a template for the next round of synthesis and screening. The polysome display method can be coupled to the phage display system. Following several rounds of screening, cDNA from the enriched pool of polysomes was cloned into a phagemid vector. This vector serves as both a peptide expression vector, displaying peptides fused to the coat proteins, and as a DNA sequencing vector for peptide identification. By expressing the polysome-derived peptides on phage, one can either continue the affinity selection procedure in this format or assay the peptides on individual clones for binding activity in a phage ELISA, or for binding specificity in a completion phage ELISA (Barret, et al. (1992) *Anal. Biochem* 204,357-364). To identify the sequences of the active peptides one sequences the DNA produced by the phagemid host.

Secondary Screening of Polypeptides and Analogs

The high through-put assays described above can be followed by secondary screens in order to identify further biological activities which will, e.g., allow one skilled in the art to differentiate agonists from antagonists. The type of a secondary screen used will depend on the desired activity that needs to be tested. For example, an assay can be developed in which the ability to inhibit an interaction between a protein of interest and its respective ligand can be used to identify antagonists from a group of peptide fragments isolated though one of the primary screens described above.

Therefore, methods for generating fragments and analogs and testing them for activity are known in the art. Once the core sequence of interest is identified, it is routine for one skilled in the art to obtain analogs and fragments.

Peptide Mimetics of *H. pylori* Polypeptides

The invention also provides for reduction of the protein binding domains of the subject *H. pylori* polypeptides to generate mimetics, e.g. peptide or non-peptide agents. The peptide mimetics are able to disrupt binding of a polypeptide to its counter ligand, e.g., in the case of an *H. pylori* polypeptide binding to a naturally occurring ligand. The critical residues of a subject *H. pylori* polypeptide which are involved in molecular recognition of a polypeptide can be determined and used to generate *H. pylori*-derived peptidomimetics which competitively or noncompetitively inhibit binding of the *H. pylori* polypeptide with an interacting polypeptide (see, for example, European patent applications EP-412,762A and EP-B31,080A).

For example, scanning mutagenesis can be used to map the amino acid residues of a particular *H. pylori* polypeptide involved in binding an interacting polypeptide, peptidomimetic compounds (e.g. diazepine or isoquinoline derivatives) can be generated which mimic those residues in binding to an interacting polypeptide, and which therefore can inhibit binding of an *H. pylori* polypeptide to an interacting polypeptide and thereby interfere with the function of *H. pylori* polypeptide. For instance, non-hydrolyzable peptide analogs of such residues can be generated using benzodiazepine (e.g., see Freidinger et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), azepine (e.g., see Huffman et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), substituted gama lactam rings (Garvey et al. in *Peptides: Chemistry and Biology*, G.R. Marshall ed., ESCOM Publisher: Leiden, Netherlands, 1988), keto-methylene pseudopeptides (Ewenson et al. (1986) *J Med Chem* 29:295; and Ewenson et al. in *Peptides: Structure and Function* (Proceedings of the 9th American Peptide Symposium) Pierce Chemical Co. Rockland, IL, 1985), β-turn dipeptide cores (Nagai et al. (1985) *Tetrahedron Lett* 26:647; and Sato et al. (1986) *J Chem Soc Perkin Trans* 1:1231), and β-aminoalcohols (Gordon et al. (1985) *Biochem Biophys Res Commun*126:419; and Dann et al. (1986) *Biochem Biophys Res Commun* 134:71).

VI. Vaccine Formulations for *H. pylori* Nucleic Acids and Polypeptides

This invention also features vaccine compositions or formulations (used interchangeably herein) for protection against infection by *H. pylori* or for treatment of *H. pylori* infection. As used herein, the term "treatment of *H. pylori* infection" refers to

therapeutic treatment of an existing or established *H. pylori* infection. The terms "protection against *H. pylori* infection" or "prophylactic treatment" refer to the use of *H. pylori* vaccine formulation for reducing the risk of or preventing an infection in a subject at risk for *H. pylori* infection. In one embodiment, the vaccine compositions contain one or more immunogenic components, such as a surface protein, from *H. pylori,* or portion thereof, and a pharmaceutically acceptable carrier. For example, in one embodiment, the vaccine formulations of the invention contain at least one or combination of *H. pylori* polypeptides or fragments thereof, from same or different *H. pylori* antigens. Nucleic acids and *H. pylori* polypeptides for use in the vaccine formulations of the invention include the nucleic acids and polypeptides set forth in the Sequence Listing, preferably those *H. pylori* nucleic acids that encode surface proteins and surface proteins or fragments thereof. For example, a preferred nucleic acid and *H. pylori* polypeptide for use in a vaccine composition of the invention is selected from the group of nucleic acids which encode cell envelope proteins and *H. pylori* cell envelope proteins as set forth in Table 1. However, any nucleic acid encoding an immunogenic *H. pylori* protein and *H. pylori* polypetide, or portion thereof, can be used in the present invention. These vaccines have therapeutic and/or prophylactic utilities.

One aspect of the invention provides a vaccine composition for protection against infection by *H. pylori* which contains at least one immunogenic fragment of an *H. pylori* protein and a pharmaceutically acceptable carrier. Preferred fragments include peptides of at least about 10 amino acid residues in length, preferably about 10-20 amino acid residues in length, and more preferably about 12-16 amino acid residues in length.

Immunogenic components of the invention can be obtained, for example, by screening polypeptides recombinantly produced from the corresponding fragment of the nucleic acid encoding the full-length *H. pylori* protein. In addition, fragments can be chemically synthesized using techniques known in the art such as conventional Merrifield solid phase f-Moc or t-Boc chemistry.

In one embodiment, immunogenic components are identified by the ability of the peptide to stimulate T cells. Peptides which stimulate T cells, as determined by, for example, T cell proliferation or cytokine secretion are defined herein as comprising at least one T cell epitope. T cell epitopes are believed to be involved in initiation and perpetuation of the immune response to the protein allergen which is responsible for the clinical symptoms of allergy. These T cell epitopes are thought to trigger early events at the level of the T helper cell by binding to an appropriate HLA molecule on the surface of an antigen presenting cell, thereby stimulating the T cell subpopulation with the relevant T cell receptor for the epitope. These events lead to T cell proliferation, lymphokine secretion, local inflammatory reactions, recruitment of additional immune

cells to the site of antigen/T cell interaction, and activation of the B cell cascade, leading to the production of antibodies. A T cell epitope is the basic element, or smallest unit of recognition by a T cell receptor, where the epitope comprises amino acids essential to receptor recognition (e.g., approximately 6 or 7 amino acid residues). Amino acid sequences which mimic those of the T cell epitopes are within the scope of this invention.

In another embodiment, immunogenic components of the invention are identified through genomic vaccination. The basic protocol is based on the idea that expression libraries consisting of all or parts of a pathogen genome, e.g., an *H. pylori* genome, can confer protection when used to genetically immunize a host. This expression library immunization (ELI) is analogous to expression cloning and involves reducing a genomic expression library of a pathogen, e.g., *H. pylori*, into plasmids that can act as genetic vaccines. The plasmids can also be designed to encode genetic adjuvants which can dramatically stimulate the humoral response. These genetic adjuvants can be introduced at remote sites and act as well extracelluraly as intracellularly.

This is a new approach to vaccine production that has many of the advantages of live/attenuated pathogens but no risk of infection. An expression library of pathogen DNA is used to immunize a host thereby producing the effects of antigen presentation of a live vaccine without the risk. For example, in the present invention, random fragments from the *H. pylori* genome or from cosmid or plasmid clones, as well as PCR products from genes identified by genomic sequencing, can be used to immunize a host. The feasibility of this approach has been demonstrated with *Mycoplasma pulmonis* (Barry et al., *Nature* 377:632-635, 1995), where even partial expression libraries of *Mycoplasma pulmonis*, a natural pathogen in rodents, provided protection against challenge from the pathogen.

ELI is a technique that allows for production of a non-infectious multipartite vaccine, even when little is known about pathogen's biology, because ELI uses the immune system to screen candidate genes. Once isolated, these genes can be used as genetic vaccines or for development of recombinant protein vaccines. Thus, ELI allows for production of vaccines in a systematic, largely mechanized fashion.

Screening immunogenic components can be accomplished using one or more of several different assays. For example, *in vitro*, peptide T cell stimulatory activity is assayed by contacting a peptide known or suspected of being immunogenic with an antigen presenting cell which presents appropriate MHC molecules in a T cell culture. Presentation of an immunogenic *H. pylori* peptide in association with appropriate MHC molecules to T cells in conjunction with the necessary costimulation has the effect of transmitting a signal to the T cell that induces the production of increased levels of

cytokines, particularly of interleukin-2 and interleukin-4. The culture supernatant can be obtained and assayed for interleukin-2 or other known cytokines. For example, any one of several conventional assays for interleukin-2 can be employed, such as the assay described in *Proc. Natl. Acad. Sci USA*, 86: 1333 (1989) the pertinent portions of which are incorporated herein by reference. A kit for an assay for the production of interferon is also available from Genzyme Corporation (Cambridge, MA).

Alternatively, a common assay for T cell proliferation entails measuring tritiated thymidine incorporation. The proliferation of T cells can be measured *in vitro* by determining the amount of ^{3}H-labeled thymidine incorporated into the replicating DNA of cultured cells. Therefore, the rate of DNA synthesis and, in turn, the rate of cell division can be quantified.

Vaccine compositions or formulations of the invention containing one or more immunogenic components (e.g., *H. pylori* polypeptide or fragment thereof or nucleic acid encoding an *H. pylori* polypeptide or fragment thereof) preferably include a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable pharmaceutically acceptable carriers include, for example, one or more of water, saline, phosphate buffered saline, dextrose, glycerol, ethanol and the like, as well as combinations thereof. Pharmaceutically acceptable carriers may further comprise minor amounts of auxiliary substances such as wetting or emulsifying agents, preservatives or buffers, which enhance the shelf life or effectiveness of the *H. pylori* nucleic acid or polypeptide. For vaccine formulations of the invention containing *H. pylori* polypeptides, the polypeptide is preferably coadministered with a suitable adjuvant and/or a delivery system described herein.

It will be apparent to those of skill in the art that the therapeutically effective amount of DNA or protein of this invention will depend, *inter alia*, upon the administration schedule, the unit dose of an *H. pylori* nucleic acid or polypeptide administered, whether the protein or nucleic acid is administered in combination with other therapeutic agents, the immune status and health of the patient, and the therapeutic activity of the particular protein or nucleic acid.

Vaccine formulations are conventionally administered parenterally, e.g., by injection, either subcutaneously or intramuscularly. Methods for intramuscular immunization are described by Wolff et al. (1990) *Science* 247: 1465-1468 and by Sedegah et al. (1994) *Immunology* 91: 9866-9870. Other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Oral immunization is preferred over parenteral methods for inducing protection against

infection by *H. pylori*. Czinn et. al. (1993) *Vaccine* 11: 637-642. Oral formulations include such normally employed excipients as, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like.

In one embodiment, the vaccine formulation includes, as a pharmaceutically acceptable carrier, an adjuvant. Examples of the suitable adjuvants for use in the vaccine formulations of the invention include, but are not limited, to aluminum hydroxide; N-acetyl-muramyl--L-threonyl-D-isoglutamine (thr-MDP); N-acetyl-nor-muramyl-L-alanyl-D-isoglutamine (CGP 11637, referred to as nor-MDP); N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphos-phoryloxy)-ethylamine (CGP 19835A, referred to a MTP-PE); RIBI, which contains three components from bacteria; monophosphoryl lipid A; trehalose dimycoloate; cell wall skeleton (MPL + TDM + CWS) in a 2% squalene/Tween 80 emulsion; and cholera toxin. Others which may be used are non-toxic derivatives of cholera toxin, including its B subunit, and/or conjugates or genetically engineered fusions of the *H. pylori* polypeptide with cholera toxin or its B subunit, procholeragenoid, fungal polysaccharides, including schizophyllan, muramyl dipeptide, muramyl dipeptide derivatives, phorbol esters, labile toxin of *E. coli*, non-*H. pylori* bacterial lysates, block polymers or saponins.

In another embodiment, the vaccine formulation includes, as a pharmaceutically acceptable carrier, a delivery system. Suitable delivery systems for use in the vaccine formulations of the invention include biodegradable microcapsules or immuno-stimulating complexes (ISCOMs), cochleates, or liposomes, genetically engineered attenuated live vectors such as viruses or bacteria, and recombinant (chimeric) virus-like particles, e.g., bluetongue. In another embodiment of the invention, the vaccine formulation includes both a delivery system and an adjuvant.

Delivery systems in humans may include enteric release capsules protecting the antigen from the acidic environment of the stomach, and including *H. pylori* polypeptide in an insoluble form as fusion proteins. Suitable carriers for the vaccines of the invention are enteric coated capsules and polylactide-glycolide microspheres. Suitable diluents are 0.2 N NaHCO3 and/or saline.

Vaccines of the invention can be administered as a primary prophylactic agent in adults or in children, as a secondary prevention, after successful eradication of *H. pylori* in an infected host, or as a therapeutic agent in the aim to induce an immune response in a susceptible host to prevent infection by *H. pylori*. The vaccines of the invention are administered in amounts readily determined by persons of ordinary skill in the art. Thus, for adults a suitable dosage will be in the range of 10 μg to 10 g, preferably 10 μg

to 100 mg, for example 50 μg to 50 mg. A suitable dosage for adults will also be in the range of 5 μg to 500 mg. Similar dosage ranges will be applicable for children.

The amount of adjuvant employed will depend on the type of adjuvant used. For example, when the mucosal adjuvant is cholera toxin, it is suitably used in an amount of 5 μg to 50 μg, for example 10 μg to 35 μg. When used in the form of microcapsules, the amount used will depend on the amount employed in the matrix of the microcapsule to achieve the desired dosage. The determination of this amount is within the skill of a person of ordinary skill in the art.

Those skilled in the art will recognize that the optimal dose may be more or less depending upon the patient's body weight, disease, the route of administration, and other factors. Those skilled in the art will also recognize that appropriate dosage levels can be obtained based on results with known oral vaccines such as, for example, a vaccine based on an *E. coli* lysate (6 mg dose daily up to total of 540 mg) and with an enterotoxigenic *E. coli* purified antigen (4 doses of 1 mg) (Schulman et al., *J. Urol.* 150:917-921 (1993)); Boedecker et al., *American Gastroenterological Assoc.* 999:A-222 (1993)). The number of doses will depend upon the disease, the formulation, and efficacy data from clinical trials. Without intending any limitation as to the course of treatment, the treatment can be administered over 3 to 8 doses for a primary immunization schedule over 1 month (Boedeker, *American Gastroenterological Assoc.* 888:A-222 (1993)).

In a preferred embodiment, a vaccine composition of the invention can be based on a killed whole *E. coli* preparation with an immunogenic fragment of an *H. pylori* protein of the invention expressed on its surface or it can be based on an *E. coli* lysate, wherein the killed *E. coli* acts as a carrier or an adjuvant.

It will be apparent to those skilled in the art that some of the vaccine compositions of the invention are useful only for preventing *H. pylori* infection, some are useful only for treating *H. pylori* infection, and some are useful for both preventing and treating *H. pylori* infection. In a preferred embodiment, the vaccine composition of the invention provides protection against *H. pylori* infection by stimulating humoral and/or cell-mediated immunity against *H. pylori*. It should be understood that amelioration of any of the symptoms of *H. pylori* infection is a desirable clinical goal, including a lessening of the dosage of medication used to treat *H. pylori*-caused disease, or an increase in the production of antibodies in the serum or mucous of patients.

VII. Antibodies Reactive With *H. pylori* Polypeptides

The invention also includes antibodies specifically reactive with the subject *H. pylori* polypeptide. Anti-protein/anti-peptide antisera or monoclonal antibodies can be made by standard protocols (See, for example, *Antibodies: A Laboratory Manual* ed. by Harlow and Lane (Cold Spring Harbor Press: 1988)). A mammal such as a mouse, a hamster or rabbit can be immunized with an immunogenic form of the peptide. Techniques for conferring immunogenicity on a protein or peptide include conjugation to carriers or other techniques well known in the art. An immunogenic portion of the subject *H. pylori* polypeptide can be administered in the presence of adjuvant. The progress of immunization can be monitored by detection of antibody titers in plasma or serum. Standard ELISA or other immunoassays can be used with the immunogen as antigen to assess the levels of antibodies.

In a preferred embodiment, the subject antibodies are immunospecific for antigenic determinants of the *H. pylori* polypeptides of the invention, e.g. antigenic determinants of a polypeptide of the invention contained in the Sequence Listing, or a closely related human or non-human mammalian homolog (e.g., 90% homologous, more preferably at least 95% homologous). In yet a further preferred embodiment of the invention, the anti-*H. pylori* antibodies do not substantially cross react (i.e., react specifically) with a protein which is for example, less than 80% percent homologous to a sequence of the invention contained in the Sequence Listing. By "not substantially cross react", it is meant that the antibody has a binding affinity for a non-homologous protein which is less than 10 percent, more preferably less than 5 percent, and even more preferably less than 1 percent, of the binding affinity for a protein of the invention contained in the Sequence Listing. In a most preferred embodiment, there is no crossreactivity between bacterial and mammalian antigens.

The term antibody as used herein is intended to include fragments thereof which are also specifically reactive with *H. pylori* polypeptides. Antibodies can be fragmented using conventional techniques and the fragments screened for utility in the same manner as described above for whole antibodies. For example, $F(ab')_2$ fragments can be generated by treating antibody with pepsin. The resulting $F(ab')_2$ fragment can be treated to reduce disulfide bridges to produce Fab' fragments. The antibody of the invention is further intended to include bispecific and chimeric molecules having an anti-*H. pylori* portion.

Both monoclonal and polyclonal antibodies (Ab) directed against *H. pylori* polypeptides or *H. pylori* polypeptide variants, and antibody fragments such as Fab` and $F(ab`)_2$, can be used to block the action of *H. pylori* polypeptide and allow the study of the role of a particular *H. pylori* polypeptide of the invention in aberrant or unwanted

intracellular signaling, as well as the normal cellular function of the *H. pylori* and by microinjection of anti-*H. pylori* polypeptide antibodies of the present invention.

Antibodies which specifically bind *H. pylori* epitopes can also be used in immunohistochemical staining of tissue samples in order to evaluate the abundance and pattern of expression of *H. pylori* antigens. Anti *H. pylori* polypeptide antibodies can be used diagnostically in immuno-precipitation and immuno-blotting to detect and evaluate *H. pylori* levels in tissue or bodily fluid as part of a clinical testing procedure. Likewise, the ability to monitor *H. pylori* polypeptide levels in an individual can allow determination of the efficacy of a given treatment regimen for an individual afflicted with such a disorder. The level of an *H. pylori* polypeptide can be measured in cells found in bodily fluid, such as in urine samples or can be measured in tissue, such as produced by gastric biopsy. Diagnostic assays using anti-*H. pylori* antibodies can include, for example, immunoassays designed to aid in early diagnosis of *H. pylori* infections. The present invention can also be used as a method of detecting antibodies contained in samples from individuals infected by this bacterium using specific *H. pylori* antigens.

Another application of anti-*H. pylori* polypeptide antibodies of the invention is in the immunological screening of cDNA libraries constructed in expression vectors such as λgt11, λgt18-23, λZAP, and λORF8. Messenger libraries of this type, having coding sequences inserted in the correct reading frame and orientation, can produce fusion proteins. For instance, λgt11 will produce fusion proteins whose amino termini consist of ß-galactosidase amino acid sequences and whose carboxy termini consist of a foreign polypeptide. Antigenic epitopes of a subject *H. pylori* polypeptide can then be detected with antibodies, as, for example, reacting nitrocellulose filters lifted from infected plates with anti-*H. pylori* polypeptide antibodies. Phage, scored by this assay, can then be isolated from the infected plate. Thus, the presence of *H. pylori* gene homologs can be detected and cloned from other species, and alternate isoforms (including splicing variants) can be detected and cloned.

VIII. Kits Containing Nucleic Acids, Polypeptides or Antibodies of the Invention

The nucleic acid, polypeptides and antibodies of the invention can be combined with other reagents and articles to form kits. Kits for diagnostic purposes typically comprise the nucleic acid, polypeptides or antibodies in vials or other suitable vessels. Kits typically comprise other reagents for performing hybridization reactions, polymerase chain reactions (PCR), or for reconstitution of lyophilized components, such as aqueous media, salts, buffers, and the like. Kits may also comprise reagents for sample processing such as detergents, chaotropic salts and the like. Kits may also

comprise immobilization means such as particles, supports, wells, dipsticks and the like. Kits may also comprise labeling means such as dyes, developing reagents, radioisotopes, fluorescent agents, luminescent or chemiluminescent agents, enzymes, intercalating agents and the like. With the nucleic acid and amino acid sequence information provided herein, individuals skilled in art can readily assemble kits to serve their particular purpose. Kits further can include instructions for use.

IX. Drug Screening Assays Using *H. pylori* Polypeptides

By making available purified and recombinant *H. pylori* polypeptides, the present invention provides assays which can be used to screen for drugs which are either agonists or antagonists of the normal cellular function, in this case, of the subject *H. pylori* polypeptides, or of their role in intracellular signaling. Such inhibitors or potentiators may be useful as new therapeutic agents to combat *H. pylori* infections in humans. A variety of assay formats will suffice and, in light of the present inventions, will be comprehended by the skilled artisan.

In many drug screening programs which test libraries of compounds and natural extracts, high throughput assays are desirable in order to maximize the number of compounds surveyed in a given period of time. Assays which are performed in cell-free systems, such as may be derived with purified or semi-purified proteins, are often preferred as "primary" screens in that they can be generated to permit rapid development and relatively easy detection of an alteration in a molecular target which is mediated by a test compound. Moreover, the effects of cellular toxicity and/or bioavailability of the test compound can be generally ignored in the *in vitro* system, the assay instead being focused primarily on the effect of the drug on the molecular target as may be manifest in an alteration of binding affinity with other proteins or change in enzymatic properties of the molecular target. Accordingly, in an exemplary screening assay of the present invention, the compound of interest is contacted with an isolated and purified *H. pylori* polypeptide.

Screening assays can be constructed *in vitro* with a purified *H. pylori* polypeptide or fragment thereof, such as an *H. pylori* polypeptide having enzymatic activity, such that the activity of .he polypeptide produces a detectable reaction product. The efficacy of the compound can be assessed by generating dose response curves from data obtained using various concentrations of the test compound. Moreover, a control assay can also be performed to provide a baseline for comparison. Suitable products include those with distinctive absorption, fluorescence, or chemi-luminescence properties, for example, because detection may be easily automated. A variety of synthetic or naturally occurring compounds can be tested in the assay to identify those

which inhibit or potentiate the activity of the *H. pylori* polypeptide. Some of these active compounds may directly, or with chemical alterations to promote membrane permeability or solubility, also inhibit or potentiate the same activity (e.g., enzymatic activity) in whole, live *H. pylori* cells.

This invention is further illustrated by the following examples which should not be construed as limiting. The contents of all references and published patent applications cited throughout this application are hereby incorporated by reference.

EXEMPLIFICATION

I. Cloning and Sequencing of *H. pylori* DNA

H. pylori chromosomal DNA was isolated according to a basic DNA protocol outlined in Schleif R.F. and Wensink P.C., *Practical Methods in Molecular Biology*, p.98, Springer-Verlag, NY., 1981, with minor modifications. Briefly, cells were pelleted, resuspended in TE (10 mM Tris, 1 mM EDTA, pH 7.6) and GES lysis buffer (5.1 M guanidium thiocyanate, 0.1 M EDTA, pH 8.0, 0.5% N-laurylsarcosine) was added. Suspension was chilled and ammonium acetate (NH_4Ac) was added to final concentration of 2.0 M. DNA was extracted, first with chloroform, then with phenol-chloroform, and reextracted with chloroform. DNA was precipitated with isopropanol, washed twice with 70% EtOH, dried and resuspended in TE.

Following isolation whole genomic *H. pylori* DNA was nebulized (Bodenteich et al., *Automated DNA Sequencing and Analysis* (J.C. Venter, ed.), Academic Press, 1994) to a median size of 2000 bp. After nebulization, the DNA was concentrated and separated on a standard 1% agarose gel. Several fractions, corresponding to approximate sizes 900-1300 bp, 1300-1700 bp, 1700-2200 bp, 2200-2700 bp, were excised from the gel and purified by the GeneClean procedure (Bio101, Inc.).

The purified DNA fragments were then blunt-ended using T4 DNA polymerase. The healed DNA was then ligated to unique BstXI-linker adapters in 100-1000 fold molar excess. These linkers are complimentary to the BstXI-cut pMPX vectors, while the overhang is not self-complimentary. Therefore, the linkers will not concatemerize nor will the cut-vector religate itself easily. The linker-adopted inserts were separated from the unincorporated linkers on a 1% agarose gel and purified using GeneClean. The linker-adopted inserts were then ligated to each of the 20 pMPX vectors to construct a series of "shotgun" subclone libraries. The vectors contain an out-of-frame lacZ gene at the cloning site which becomes in-frame in the event that an adapter-dimer is cloned, allowing these to be avoided by their blue-color.

All subsequent steps were based on the multiplex DNA sequencing protocols outlined in Church G.M. and Kieffer-Higgins S., *Science* 240:185-188, 1988. Only major modifications to the protocols are highlighted. Briefly, each of the 20 vectors was then transformed into DH5α competent cells (Gibco/BRL, DH5α transformation protocol). The libraries were assessed by plating onto antibiotic plates containing ampicillin, methicillin and IPTG/Xgal. The plates were incubated overnight at 37°C. Successful transformants were then used for plating of clones and pooling into the multiplex pools. The clones were picked and pooled into 40 ml growth medium cultures. The cultures were grown overnight at 37°C. DNA was purified using the Qiagen Midi-prep kits and Tip-100 columns (Qiagen, Inc.). In this manner, 100 μg of DNA was obtained per pool. Fifteen 96-well plates of DNA were generated to obtain a 5-10 fold sequence redundancy assuming 250-300 base average read-lengths.

These purified DNA samples were then sequenced using the multiplex DNA sequencing based on chemical degradation methods (Church G.M. and Kieffer-Higgins S., *Science* 240:185-188, 1988) or by Sequithrem (Epicenter Technologies) dideoxy sequencing protocols. The sequencing reactions were electrophoresed and transferred onto nylon membranes by direct transfer electrophoresis from 40 cm gels (Richterich P. and Church G.M., *Methods in Enzymology* 218:187-222, 1993) or by electroblotting (Church, *supra*). 24 samples were run per gel. 45 successful membranes were produced by chemical sequencing and 8 were produced by dideoxy sequencing. The DNA was covalently bound to the membranes by exposure to ultraviolet light, and hybridized with labeled oligonucleotides complimentary to tag sequences on the vectors (Church, *supra*). The membranes were washed to rinse off non-specifically bound probe, and exposed to X-ray film to visualize individual sequence ladders. After autoradiography, the hybridized probe was removed by incubation at 65° C, and the hybridization cycle repeated with another tag sequence until the membrane had been probed 38 times for chemical sequencing membranes and 10 times for the dideoxy sequencing membranes. Thus, each gel produced a large number of films, each containing new sequencing information. Whenever a new blot was processed, it was initially probed for an internal standard sequence added to each of the pools.

Digital images of the films were generated using a laser-scanning densitometer (Molecular Dynamics, Sunnyvale, CA). The digitized images were processed on computer workstations (VaxStation 4000's) using the program REPLICA™ (Church et al., *Automated DNA Sequencing and Analysis* (J.C. Venter, ed.), Academic Press, 1994). Image processing included lane straightening, contrast adjustment to smooth out intensity differences, and resolution enhancement by iterative gaussian deconvolution. The sequences were then automatically picked in REPLICA™ and displayed for

interactive proofreading before being stored in a project database. The proofreading was accomplished by a quick visual scan of the film image followed by mouse clicks on the bands of the displayed image to modify the base calls. Many of the sequence errors could be detected and corrected because multiple sequence reads covering the same portion of the genomic DNA provide adequate sequence redundancy for editing. Each sequence automatically received an identification number (corresponding to microtiter plate, probe information, and lane set number). This number serves as a permanent identifier of the sequence so it is always possible to identify the original of any particular sequence without recourse to a specialized database.

Routine assembly of *H. pylori* sequences was done using the program FALCON (Church, Church et al., *Automated DNA Sequenicng and Analysis* (J.C. Venter, ed.), Academic Press, 1994). This program has proven to be fast and reliable for most sequences. The assembled contigs were displayed using a modified version of GelAssemble, developed by the Genetics Computer Group (GCG) (Devereux et al., *Nucleic Acid Res.* 12:387-95, 1984) that interacts with REPLICA™. This provided for an integrated editor that allows multiple sequence gel images to be instantaneously called up from the REPLICA™ database and displayed to allow rapid scanning of contigs and proofreading of gel traces where discrepancies occurred between different sequence reads in the assembly.

II. Identification, cloning and expression of recombinant *H. pylori* DNA sequences

To facilitate the cloning, expression and purification of membrane and secreted proteins from *H. pylori* a powerful gene expression system, the pET System (Novagen), for cloning and expression of recombinant proteins in *E. coli*, was selected. Also, a DNA sequence encoding a peptide tag, the His-Tag, was fused to the 3' end of DNA sequences of interest in order to facilitate purification of the recombinant protein products. The 3' end was selected for fusion in order to avoid alteration of any 5' terminal signal sequence. The exception to the above was ppiB, a gene cloned for use as a control in the expression studies. In this study, the sequence for *H. pylori* ppiB contains a DNA sequence encoding a His-Tag fused to the 5' end of the full length gene, because the protein product of this gene does not contain a signal sequence and is expressed as a cytosolic protein.

PCR Amplification and cloning of DNA sequences containing ORF's for membrane and secreted proteins from the J99 Strain of Helicobacter pylori.

Sequences chosen (from the list of the DNA sequences of the invention) for cloning from the J99 strain of *H. pylori* were prepared for amplification cloning by polymerase chain reaction (PCR). Synthetic oligonucleotide primers (Table 3) specific for the 5' and 3' ends of open reading frames (ORFs) were designed and purchased (GibcoBRL Life Technologies, Gaithersburg, MD, USA). All forward primers (specific for the 5' end of the sequence) were designed to include an NcoI cloning site at the extreme 5' terminus, except for HpSeq. 4821082 where NdeI was used. These primers were designed to permit initiation of protein translation at a methionine residue followed by a valine residue and the coding sequence for the remainder of the native *H. pylori* DNA sequence. An exception is *H. pylori* sequence 4821082 where the initiator methionine is immediately followed by the remainder of the native *H. pylori* DNA sequence. All reverse primers (specific for the 3' end of any *H. pylori* ORF) included a EcoRI site at the extreme 5' terminus to permit cloning of each *H. pylori* sequence into the reading frame of the pET-28b. The pET-28b vector provides sequence encoding an additional 20 carboxy-terminal amino acids (only 19 amino acids in HpSeq. 26380318 and HpSeq.14640637) including six histidine residues (at the extreme C-terminus), which comprise the His-Tag. An exception to the above, as noted earlier, is the vector construction for the ppiB gene. A synthetic oligonucleotide primer specific for the 5' end of ppiB gene encoded a BamHI site at its extreme 5' terminus and the primer for the 3' end of the ppiB gene encoded a XhoI site at its extreme 5' terminus.

TABLE 3

Oligonucleotide primers used for PCR amplification of *H. pylori* DNA sequences

Outer membrane Proteins	**Forward primer 5' to 3'**	**Reverse Primer 5' to 3'**
Protein 16225006	5'-TATACCATGGTGGG CGCTAA-3' (SEQ ID NO:195)	5'-ATGAATTCGAGTA AGGATTTTTG-3' (SEQ ID NO:196)
Protein 26054702	5'-TTAACCATGGTGA AAAGCGATA-3' (SEQ ID NO:197)	5'-TAGAATTCGCATA ACGATCAATC-3' (SEQ ID NO:198)
Protein 7116626	5'-ATATCCATGGTGA GTTTGATGA-3' (SEQ ID NO:199)	5'-ATGAATTCAATTT TTTATTTTGCCA-3' (SEQ ID NO:200)

Protein 29479681	5'-AATTCCATGGTGG GGGCTATG-3' (SEQ ID NO:201)	5'-ATGAATTCTCGAT AGCCAAAATC-3' (SEQ ID NO:202)
Protein 14640637	5'-AATTCCATGGTG CATAACTTCCATT-3' (SEQ ID NO:203)	5'-AAGAATTCTCTA GCATCCAAATGGA-3' (SEQ ID NO:204)
Periplasmic/ Secreted Proteins		
Protein 30100332	5'-ATTTCCATGGTCATG TCTCATATT-3' (SEQ ID NO:205)	5'-ATGAATTCCATC TTTTATTCCAC-3' (SEQ ID NO:206)
Protein 4721061	5'-AACCATGGTGATTT TAAGCATTGAAAG-3' (SEQ ID NO:207)	5'-AAGAATTCCAC TCAAAATTTTTTAAC AG-3' (SEQ ID NO:208)
Other Surface Proteins		
Protein 4821082	5'-GATCATCCATATGTT ATCTTCTAAT-3' (SEQ ID NO:209)	5'-TGAATTCAACCA TTTTAACCCTG-3' (SEQ ID NO:210)
Protein 978477	5'-TATACCATGGTGAA ATTTTTTCTTTTA-3' (SEQ ID NO:211)	5'-AGAATTCAATT GCGTCTTGTAAAAG-3' (SEQ ID NO:212)
Inner Membrane Protein		
Protein 26380318	5'-TATACCATGGTGAT GGACAAACTC-3' (SEQ ID NO:213)	5'-ATGAATTCCCACTT GGGGCGATA-3' (SEQ ID NO:214)
Cytoplasmic Protein		
ppi	5'-TTATGGATCCAAAC CAATTAAAACT-3' (SEQ ID NO:215)	5'-TATCTCGAGTTATA GAGAAGGGC-3' (SEQ ID NO:216)

Genomic DNA prepared from the J99 strain of *H. pylori* (ATCC #55679; deposited by Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154) was used as the source of template DNA for PCR amplification reactions (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). To amplify a DNA sequence containing an *H. pylori* ORF, genomic DNA (50 nanograms) was introduced into a reaction vial containing 2 mM $MgCl_2$, 1 micromolar synthetic oligonucleotide primers (forward and reverse primers) complementary to and flanking a defined *H. pylori* ORF, 0.2 mM of each deoxynucleotide triphosphate; dATP, dGTP, dCTP, dTTP and 2.5 units of heat stable

DNA polymerase (Amplitaq, Roche Molecular Systems, Inc., Branchburg, NJ, USA) in a final volume of 100 microliters. The following thermal cycling conditions were used to obtain amplified DNA products for each ORF using a Perkin Elmer Cetus/ GeneAmp PCR System 9600 thermal cycler:

Protein 26054702, Protein 7116626, Protein 29479681, Protein 30100332, and Protein 4821082;
Denaturation at 94°C for 2 min,
2 cycles at 94°C for 15 sec, 30°C for 15 sec and 72°C for 1.5 min
23 cycles at 94°C for 15 sec, 55°C for 15 sec and 72°C for 1.5 min
Reactions were concluded at 72°C for 6 minutes.

Protein 16225006;
Denaturation at 94°C for 2 min,
25 cycles at 95°C for 15 sec, 55°C for 15 sec and 72°C for 1.5 min
Reaction was concluded at 72°C for 6 minutes.

Protein 4721061;
Denaturation at 94°C for 2 min,
2 cycles at 94°C for 15 sec, 36°C for 15 sec and 72°C for 1.5 min
23 cycles at 94°C for 15 sec, 60°C for 15 sec and 72°C for 1.5 min
Reactions were concluded at 72°C for 6 minutes.

Protein 26380318;
Denaturation at 94°C for 2 min,
2 cycles at 94°C for 15 sec, 38°C for 15 sec and 72°C for 1.5 min
23 cycles at 94°C for 15 sec, 62°C for 15 sec and 72°C for 1.5 min
Reactions were concluded at 72°C for 6 minutes.

Protein 14640637;
Denaturation at 94°C for 2 min,
2 cycles at 94°C for 15 sec, 33°C for 15 sec and 72°C for 1.5 min
30 cycles at 94°C for 15 sec, 55°C for 15 sec and 72°C for 1.5 min
Reactions were concluded at 72°C for 6 minutes.

Conditions for amplification of *H. pylori* ppiB;
Denaturation at 94°C for 2 min,
2 cycles at 94°C for 15 sec, 32°C for 15 sec and 72°C for 1.5 min
25 cycles at 94°C for 15 sec, 56°C for 15 sec and 72°C for 1.5 min
Reactions were concluded at 72°C for 6 minutes

Upon completion of thermal cycling reactions, each sample of amplified DNA was washed and purified using the Qiaquick Spin PCR purification kit (Qiagen, Gaithersburg, MD, USA). All amplified DNA samples were subjected to digestion with the restriction endonucleases, NcoI and EcoRI (New England BioLabs, Beverly, MA, USA), or in the case of HpSeq. 4821082 (SEQ ID NO: 1309), with NdeI and EcoRI (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). DNA samples were then subjected to electrophoresis on 1.0 % NuSeive (FMC BioProducts, Rockland, ME USA) agarose gels. DNA was visualized by exposure to ethidium bromide and long wave uv irradiation. DNA contained in slices isolated from the agarose gel was purified using the Bio 101 GeneClean Kit protocol (Bio 101 Vista, CA, USA).

Cloning of H. pylori DNA sequences into the pET-28b prokaryotic expression vector.

The pET-28b vector was prepared for cloning by digestion with NcoI and EcoRI, or in the case of *H. pylori* protein 4821082 with NdeI and EcoRI (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). In the case of cloning ppiB, the pET-28a vector, which encodes a His-Tag that can be fused to the 5' end of an inserted gene, was used and the cloning site prepared for cloning with the ppiB gene by digestion with BamHI and XhoI restriction endonucleases.

Following digestion, DNA inserts were cloned (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994) into the previously digested pET-28b expression vector, except for the amplified insert for ppiB, which was cloned into the pET-28a expression vector. Products of the ligation reaction were then used to transform the BL21 strain of *E. coli* (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994) as described below.

Transformation of competent bacteria with recombinant plasmids

Competent bacteria, *E coli* strain BL21 or *E. coli* strain BL21(DE3), were transformed with recombinant pET expression plasmids carrying the cloned *H. pylori* sequences according to standard methods (Current Protocols in Molecular, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). Briefly, 1 microliter of ligation reaction

was mixed with 50 microliters of electrocompetent cells and subjected to a high voltage pulse, after which, samples were incubated in 0.45 milliliters SOC medium (0.5% yeast extract, 2.0 % tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4 and 20, mM glucose) at 37°C with shaking for 1 hour. Samples were then spread on LB agar plates containing 25 microgram/ml kanamycin sulfate for growth overnight. Transformed colonies of BL21 were then picked and analyzed to evaluate cloned inserts as described below.

Identification of recombinant pET expression plasmids carrying H. pylori sequences

Individual BL21 clones transformed with recombinant pET-28b-H.pylori ORFs were analyzed by PCR amplification of the cloned inserts using the same forward and reverse primers, specific for each *H. pylori* sequence, that were used in the original PCR amplification cloning reactions. Successful amplification verified the integration of the *H. pylori* sequences in the expression vector (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994).

Isolation and Preparation of plasmid DNA from BL21 transformants

Individual clones of recombinant pET-28b vectors carrying properly cloned *H. pylori* ORFs were picked and incubated in 5 mls of LB broth plus 25 microgram/ml kanamycin sulfate overnight. The following day plasmid DNA was isolated and purified using the Qiagen plasmid purification protocol (Qiagen Inc., Chatsworth, CA, USA).

Expression of recombinant H. pylori sequences in E. coli

The pET vector can be propagated in any *E. coli* K-12 strain e.g. HMS174, HB101, JM109, DH5, etc. for the purpose of cloning or plasmid preparation. Hosts for expression include *E. coli* strains containing a chromosomal copy of the gene for T7 RNA polymerase. These hosts are lysogens of bacteriophage DE3, a lambda derivative that carries the lacI gene, the lacUV5 promoter and the gene for T7 RNA polymerase. T7 RNA polymerase is induced by addition of isopropyl-B-D-thiogalactoside (IPTG), and the T7 RNA polymerase transcribes any target plasmid, such as pET-28b, carrying a T7 promoter and a gene of interest. Strains used include: BL21(DE3) (Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990) Meth. Enzymol. 185, 60-89).

To express recombinant *H. pylori* sequences, 50 nanograms of plasmid DNA isolated as described above was used to transform competent BL21(DE3) bacteria as described above (provided by Novagen as part of the pET expression system kit). The lacZ gene (beta-galactosidase) was expressed in the pET-System as described for the *H.*

pylori recombinant constructions. Transformed cells were cultured in SOC medium for 1 hour, and the culture was then plated on LB plates containing 25 micrograms/ml kanamycin sulfate. The following day, bacterial colonies were pooled and grown in LB medium containing kanamycin sulfate (25 micrograms/ml) to an optical density at 600 nM of 0.5 to 1.0 O.D. units, at which point, 1 millimolar IPTG was added to the culture for 3 hours to induce gene expression of the *H. pylori* recombinant DNA constructions.

After induction of gene expression with IPTG, bacteria were pelleted by centrifugation in a Sorvall RC-3B centrifuge at 3500 x g for 15 minutes at 4°C. Pellets were resuspended in 50 milliliters of cold 10 mM Tris-HCl, pH 8.0, 0.1 M NaCl and 0.1 mM EDTA (STE buffer). Cells were then centrifuged at 2000 x g for 20 min at 4°C. Wet pellets were weighed and frozen at -80°C until ready for protein purification.

III. Purification of recombinant proteins from *E. coli*

Analytical Methods

The concentrations of purified protein preparations were quantified spectrophotometrically using absorbance coefficients calculated from amino acid content (Perkins, S.J. 1986 Eur. J. Biochem. 157, 169-180). Protein concentrations were also measured by the method of Bradford, M.M. (1976) Anal. Biochem. 72, 248-254, and Lowry, O.H., Rosebrough, N., Farr, A.L. & Randall, R.J. (1951) J. Biol. Chem. 193, pages 265-275, using bovine serum albumin as a standard.

SDS-polyacrylamide gels (12% or 4.0 to 25 % acrylamide gradient gels) were purchased from BioRad (Hercules, CA, USA), and stained with Coomassie blue. Molecular weight markers included rabbit skeletal muscle myosin (200 kDa), *E. coli* (-galactosidase (116 kDa), rabbit muscle phosphorylase B (97.4 kDa), bovine serum albumin (66.2 kDa), ovalbumin (45 kDa), bovine carbonic anhydrase (31 kDa), soybean trypsin inhibitor (21.5 kDa), egg white lysozyme (14.4 kDa) and bovine aprotinin (6.5 kDa).

1. Purification of soluble proteins

All steps were carried out at 4°C. Frozen cells were thawed, resuspended in 5 volumes of lysis buffer (20 mM Tris, pH 7.9, 0.5 M NaCl, 5 mM imidazole with 10% glycerol, 0.1 % 2-mercaptoethanol, 200 µg/ ml lysozyme, 1 mM phenylmethylsulfonyl fluoride (PMSF), and 10 ug/ml each of leupeptin, aprotinin, pepstatin, L-1-chloro-3-[4-tosylamido]-7-amino-2-heptanone (TLCK), L-1-chloro-3-[4-tosylamido]-4-phenyl-2-butanone (TPCK), and soybean trypsin inhibitor, and ruptured by several passages through a small volume microfluidizer (Model M-110S, Microfluidics International

Corporation, Newton, MA). The resultant homogenate was made 0.1 % Brij 35, and centrifuged at 100,000 x g for 1 hour to yield a clear supernatant (crude extract).

Following filtration through a 0.8 μm Supor filter (Gelman Sciences, FRG) the crude extract was loaded directly onto a Ni^{2+}- nitrilotriacetate-agarose (NTA) with a 5 milliliter bed volume (Hochuli, E., Dbeli, H., and Schacheer, A. (1987) J. Chromatography 411, 177-184) pre-equilibrated in lysis buffer containing 10 % glycerol, 0.1 % Brij 35 and 1 mM PMSF. The column was washed with 250 ml (50 bed volumes) of lysis buffer containing 10 % glycerol, 0.1 % Brij 35, and was eluted with sequential steps of lysis buffer containing 10 % glycerol, 0.05 % Brij 35, 1 mM PMSF, and 20, 100, 200, and 500 mM imidazole in succession. Fractions were monitored by absorbance at OD_{280} nm, and peak fractions were analyzed by SDS-PAGE. Fractions containing the recombinant protein eluted at 100 mM imidazole.

Recombinant protein 14640637 and proteins, beta-galactosidase (lacZ) and peptidyl-prolyl cis-trans isomerase (ppiB)

Fractions containing the recombinant proteins from the Ni^{2+}-NTA-agarose columns were pooled and then concentrated to approximately 5 ml by centrifugal filtration (Centriprep-10, Amicon, MA), and loaded directly onto a 180-ml column (1.6 X 91 cm) of Sephacryl S-100 HR gel filtration medium equilibrated in Buffer A (10 mM Hepes, pH 7.5, 150 mM NaCl, 0.1 mM EGTA) and run in Buffer A at 18 ml/h. Fractions containing the recombinant protein were identified by absorbance at 280 nm and analyzed by SDS-PAGE. Fractions were pooled and concentrated by centrifugal filtration.

Recombinant protein 7116626

Fractions containing the recombinant protein from the Ni^{2+} -NTA-agarose column were pooled and dialyzed overnight against 1 liter of dialysis buffer (10 mM MOPS, pH 6.5, 50 mM NaCl, 0.1 mM EGTA, 0.02% Brij 35 and 1 mM PMSF). In the morning, a fine white precipitate was removed by centrifugation and the resulting supernatant was loaded onto an 8 ml (8 x 75 mm) MonoS high performance liquid chromatography column (Pharmacia Biotechnology, Inc., Piscataway, NJ, USA) equilibrated in buffer B (10 mM MOPS, pH 6.5, 0.1 mM EGTA) containing 50 mM NaCl. The column was washed with 10 bed volumes of buffer B containing 50 mM NaCl, and developed with a 50-ml linear gradient of increasing NaCl (50 to 500 mM). Recombinant protein 7116626 eluted as a sharp peak at 300 mM NaCl.

2. Purification of insoluble proteins from inclusion bodies

The following steps were carried out at 4°C. Cell pellets were resuspended in lysis buffer with 10% glycerol 200 µg/ ml lysozyme, 5 mM EDTA, 1mM PMSF and 0.1 % -mercaptoethanol. After passage through the cell disrupter, the resulting homogenate was made 0.2 % deoxycholate, stirred 10 minutes, then centrifuged at 20,000 x g, for 30 min. The pellets were washed with lysis buffer containing 10 % glycerol, 10 mM EDTA, 1% Triton X-100, 1 mM PMSF and 0.1% -mercaptoethanol, followed by several washes with lysis buffer containing 1 M urea, 1 mM PMSF and 0.1 % 2-mercaptoethanol. The resulting white pellet was composed primarily of inclusion bodies, free of unbroken cells and membranous materials.

Recombinant proteins 26054702, 16225006, 30100332, 4721061

The following steps were carried out at room temperature. Purified inclusion bodies were dissolved in 20 ml 8.0 M urea in lysis buffer with 1 mM PMSF and 0.1 % 2-mercaptoethanol, and incubated at room temperature for 1 hour. Materials that did not dissolve were removed by centrifugation. The clear supernatant was filtered, then loaded onto a Ni^{2+} -NTA agarose column pre-equilibrated in 8.0 M urea in Lysis Buffer. The column was washed with 250 ml (50 bed volumes) of lysis buffer containing 8 M urea, 1.0 mM PMSF and 0.1 % 2-mercaptoethanol, and developed with sequential steps of lysis buffer containing 8M urea, 1 mM PMSF, 0.1 % 2-mercaptoethanol and 20, 100, 200, and 500 mM imidazole in succession. Fractions were monitored by absorbance at OD_{280} nm, and peak fractions were analyzed by SDS-PAGE. Fractions containing the recombinant protein eluted at 100 mM imidazole.

Recombinant proteins 29479681, 26380318

The pellet containing the inclusion bodies was solubilized in buffer B containing 8 M urea, 1 mM PMSF and 0.1 % 2-mercaptoethanol, and incubated for 1 hour at room temperature. Insoluble materials were removed by centrifugation at 20,000 x g for 30 min, and the cleared supernatant was loaded onto a 15 ml (1.6 x 7.5 cm) SP-Sepharose column pre-equilibrated in buffer B, 6 M urea, 1 mM PMSF, 0.1 % 2-mercaptoethanol. After washing the column with 10 bed volumes, the column was developed with a linear gradient from 0 to 500 mM NaCl.

Dialysis and concentration of protein samples

Urea was removed slowly from the protein samples by dialysis against Tris-buffered saline (TBS; 10 mM Tris pH 8.0, 150 mM NaCl) containing 0.5 % deoxycholate (DOC) with sequential reduction in urea concentration as follows; 6M,

4M, 3M, 2M, 1M, 0.5 M and finally TBS without any urea. Each dialysis step was conducted for a minimum of 4 hours at room temperature.

After dialysis, samples were concentrated by pressure filtration using Amicon stirred-cells. Protein concentrations were measured using the methods of Perkins (1986 Eur. J. Biochem. 157, 169-180), Bradford ((1976) Anal. Biochem. 72, 248-254) and Lowry ((1951) J. Biol. Chem. 193, pages 265-275).

The recombinant proteins purified by the methods described above are summarized in Table 4 below.

TABLE 4

J99 Sequence Identifier	Homolog identified by Blast	Gene symbol of Homolog	Bacterial cell fraction used to purify recombinant proteins	Method of purification	Relative MW on SDS-PAGE gel	Final concentration of purified protein	Composition of buffer
Outer Membrane Proteins							
16225006	P28635	**YEAC**	Inclusion bodies	His-Tag	18 kDa	5 mg/ml	B
26054702	P15929	**flgH**	Inclusion bodies	His-Tag	37 kDa	1.18 mg/ml	B
						----	as dry pellet
7116626	P26093	**e(P4)**	Soluble fraction	His-Tag	29 kDa	0.8 mg/ml	A
						1.85 mg/ml	C
29479681	P13036	**fecA**	Inclusions bodies	SP-Sepharose	23 kDa	2.36 mg/ml	B
						0.5 mg ml	B
						----	as dry pellet
14640637	P16665	**TPF1**	Soluble fraction	His-Tag	17 kDa	2.4 mg/ml	A
				gel filtration S100 HR			
Periplasmic/Secreted Protein							
3[illegible]032	P23847	**dppA**	Inclusion bodies	His-Tag	11 kDa	2.88 mg/ml	B
4721061	P36175	**GCP**	Inclusion bodies	His-Tag	38 kDa	2.8 mg/ml	B

Other Surface Proteins

4821082	P08089	M protein	Inclusion bodies	His-Tag	20 kDa	1.16 mg/ml	B
978477	L28919	FBP54	Inclusion bodies	SP-Sepharose	44 kDa	2.56 mg/ml	B
						0.3 mg/ml	B

Inner Membrane Proteins

26380318	P15933	fliG	Inclusion bodies	SP-Sepharose	11 kDa	22 mg/ml	B

Control Proteins with His-Tag

	P00722	lacZ	Soluble fraction	His-Tag	116 kDa	10 mg/ml	A
				gel filtration S200 HR			
		ppiB	Soluble fraction	His-Tag	21 kDa	4.4 mg/ml	A
				gel filtration S100 HR			
Buffer compositions:							
A=10 mM Hepes pH 7.5, 150 mM NaCl, 0.1 mM EGTA							
B= 10 mM Tris pH 8.0, 150 mM NaCl, 0.5 % DOC							
C= 10 mM MOPS pH 6.5, 300 mM NaCl, 0.1 EGTA							

IV. Analysis of *H. pylori* proteins as Vaccine candidates

To analyze *H. pylori* proteins for use in the vaccine formulations of the invention, several *H. pylori* proteins were expressed, characterized immunologically and tested in animal efficacy studies as outlined below. Specifically, the immunomodulatory effects of *H. pylori* proteins were investigated in a mouse/*H. pylori* model which mimics the human *H. pylori* infection in humans. In these studies, the effect of oral immunization of selected *H. pylori* polypeptides in *H. pylori* infected mice was determined.

Identification, cloning and expression of recombinant Helicobacter pylori sequences.

To facilitate the cloning, expression and purification of membrane and/or secreted proteins from *H. pylori*, the pET gene expression system (Novagen), for cloning and expression of recombinant proteins in *Escherichia coli* was selected. Further, for proteins that have a signal sequence at their amino-terminal end, a DNA sequence encoding a peptide tag (His-tag) was fused to the 5' end of the *H. pylori* DNA sequences of interest in order to facilitate purification of the recombinant protein products.

PCR amplification and cloning of DNA sequences containing ORFs for membrane and secreted proteins from the J99 strain of Helicobacter pylori.

The sequences selected (from the list of the DNA sequences of the invention) for cloning from *H. pylori* strain J99 were prepared for amplification cloning by the polymerase chain reaction (PCR). All of the selected sequences encode for outer membrane *H. pylori* proteins, with vac9 (SEQ ID NO:125), vac10 (SEQ ID NO:147), vac22 (SEQ ID NO:121) and vac41 (SEQ ID NO:176) sequences all sharing a terminal phenylalanine residue. Likewise, the vac32 (SEQ ID NO:108), vac36 (SEQ ID NO:149) and vac37 (SEQ ID NO:139) sequences all share a terminal phenylalanine residue and a tyrosine cluster at the C-terminus. Synthetic oligonucleotide primers for each ORF of interest (Table 5) specific for the predicted mature 5' end of the ORF and downstream (3') of the predicted translational termination codon were designed and purchased (GibcoBRL Life Technologies, Gaithersburg, MD, USA). All forward primers (specific for the 5' terminus of the region of ORF of interest) were designed to include a *Bam*III restriction site followed by a *Nde*I restriction site. These primers were designed to permit the initiation of protein translation at a methionine residue encoding within the *Nde*I restriction site sequence (in the case of producing a non His-tagged recombinant protein) or to fuse in frame with the DNA sequence encoding the His-tag (for producing His-tagged recombinant protein), followed by the coding sequence for the remainder of the native *H. pylori* DNA. All reverse oligonucleotide primers (specific for downstream (3') of the predicted translational termination codon of the ORF) were designed to include an *Eco*RI restriction site at the 5' terminus. This combination of primers would enable each ORF of interest to be cloned into pET28b (to produce a His-tagged recombinant protein) or pET30a (to produce a non His-tagged or native recombinant protein). The pET28b vector provides sequence encoding an additional 20 amino-terminal amino acids (plus the methionine in the *Nde*I restriction site) including a stretch of six histidine residues which makes up the His-tag.

Genomic DNA prepared from *H. pylori* strain J99 (ATCC 55679) was used as the source of template DNA for the PCR amplification reactions (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). To amplify a DNA sequence containing a specific *H. pylori* ORF, genomic DNA (50 nanograms) was introduced into a reaction tube containing 200 nanograms of both the forward and reverse synthetic oligonucleotide primer specific for the ORF of interest, and 45 microliters of PCR SuperMix purchased (GibcoBRL Life Technologies, Gaithersburg, MD, USA) in a total of 50 microliters. The PCR SuperMix is supplied in 1.1X concentrations and contains 22mM Tris-HCl (pH 8.4), 55mM KCl, 1.65 mM

$MgCl_2$, 220 micromolar of each dATP, dCTP, dGTP and dTTP, 22 units recombinant *Taq* polymerase/ml and stabilizers. The following thermal cycling conditions were used to obtain amplified DNA products for each ORF using a Perkin Elmer Cetus/Gene Amp PCR System thermal cycler.

Table 5: Oligonucleotide primers

Gene	Forward primer	Reverse primer
vac9 (nt SEQ ID NO:28) (aa SEQ ID NO: 125)	CGCGGATCCATATGGCTGAAA AAACGCCTTTTTTTAAAACTAA AAACCAC (SEQ ID NO: 257)	CCGGAATTCATCAGTATTCAA TGGGAATAAAGCC (SEQ ID NO: 258)
vac10 (nt SEQ ID NO:50) (aa SEQ ID NO: 147)	CGCGGATCCATATGAAAGAAG AAGAAAAAGAAGAAAAAAAG ACAGAAAGG (SEQ ID NO: 259)	CCGGAATTCGCTTAAAAGAAA ATAGTCCCCCAAACGC (SEQ ID NO: 260)
vac22 (nt SEQ ID NO:24) (aa SEQ ID NO: 121)	CGCCGGATCCATATGAAAGAG GTCATTCCACCCCTTCAACCCC (SEQ ID NO: 261)	CCGGAATTCATATAAATATCA TATAGGCAGAAAAAC (SEQ ID NO: 262)
vac32 (nt SEQ ID NO:11) (aa SEQ ID NO: 108)	CGCGGATCCATATGGAGGCAG AGCTTGATGAAAAATC (SEQ ID NO: 263)	CCGGAATTCGATTGATTTTGTC AAATCTAAAATCCC (SEQ ID NO: 264)
vac36 (hop B) (nt SEQ ID NO:52) (aa SEQ ID NO:149)	TATTATACATATGGAAGAAGA TGGG (SEQ ID NO: 265)	TAATCTCGAGTTTAGAAGGCG TA (SEQ ID NO: 266)

vac37 (i-hop) (nt SEQ ID NO:42) (aa SEQ ID NO: 139)	TTATATTCATATGGAAGACGATGGC (SEQ ID NO: 267)	AATTCTCGAGCCTCTTTATAAGCC (SEQ ID NO: 268)
vac41 (nt SEQ ID NO:79) (aa SEQ ID NO: 176)	CGCGGATCCATATGGTAGAAGCCTTTCAAAAACACCAAAAAGACGG (SEQ ID NO: 269)	CCGGAATTCGGAGCCAATAGGGAGCTAAAGCC (SEQ ID NO: 270)

Sequences for Vac32, Vac9 and Vac22

Denaturation at 94°C for 30 sec

35 cycles at 94°C for 15 sec, 55°C for 15 sec, and 72°C for 1.5 min

Reactions were concluded at 72°C for 8 minutes

Sequences for Vac10 and Vac41

Denaturation at 94°C for 30 sec

35 cycles at 94°C for 15 sec, 55°C for 15 sec, and 72°C for 2.5 min

Reactions were concluded at 72°C for 8 minutes

Sequences for Vac36 and Vac37

Denaturation at

2 cycles at 94°C for 15 sec, 30°C for 15 sec, and 72°C for 1.5 min

23 cycles at 94°C for 15 sec, 55°C for 15 sec, and 72°C for 1.5 min

Reactions were concluded at 72°C for 6 minutes

Upon completion of the thermal cycling reactions, each sample of amplified DNA was subjected to electrophoresis on 1.0% agarose gels. The DNA was visualized by exposure to ethidium bromide and long wave UV irradiation, and cut out in gel slices. DNA was purified using the Wizard PCR Preps Kit (Promega Corp., Madison, WI, USA), and then subjected to digestion with *Bam*HI and *Eco*RI (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). The digested PCR amplicon was then re-electrophoresed and purified as before.

Ligation of H. pylori DNA sequences into cloning vectors

The pOK12 vector (J. Vieira and J. Messing, Gene 100:189-194, 1991) was prepared for cloning for digestion with *Bam*HI and *Eco*RI in the case of Vac9, 10, 22, 31 and 32, whereas the pSU21 vector (B. Bartolome *et al.*, Gene 102:75-78, 1991) was prepared for cloning by digestion with *Bam*HI and *Eco*RI in the case of Vac 41 (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). The vectors were subjected to electrophoresis on 1.0% agarose gels and purified using the Wizard PCR Preps kit (Promega Corp., Madison, WI, USA). Following ligation of the purified, digested vector and the purified, digested amplified *H. pylori* ORF, the products of the ligation reaction were transformed into *E. coli* JM109 competent cells according to standard methods (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). Individual bacterial colonies were screened for those containing the correct recombinant plasmids by incubating in LB broth overnight (plus 25ug/ml kanamycin sulfate for the pOK12 based plasmids or 25ug/ml chloramphenicol for the pSU21 based plasmids) followed by plasmid DNA preparation using the Magic Minipreps system (Promega Corp., Madison, WI, USA), and then analyzed by restriction digestion (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994).

Cloning of H. pylori DNA sequences into the pET28b and pET30a prokaryotic expression vectors

Both the pET28b and pET30a expression vectors were prepared for cloning by digestion with *Nde*I and *Eco*RI (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). The *H. pylori* DNA sequences were removed from pOK12 (Vac9,10,23,31 and 32) or pSU21 (Vac41) plasmid backbones by digestion with N*de*I and *Eco*RI (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). The pET28b, pET30a and H. *pylori* DNA sequences were all electrophoresed on a 1% agarose gel and purified using the Wizard PCR Preps kit (Promega Corp., Madison WI, USA). Following ligation of the purified, digested expression vector and the purified, digest *H. pylori* DNA sequences, the products of the ligation reaction were transformed into *E. coli* JM109 competen. cells (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). Individual bacterial colonies were screened for those containing the correct recombinant plasmids by preparing plasmid DNA as described above followed by analysis by restriction digestion profiles and DNA sequencing (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). These recombinant plasmids were then used to transform specific *E. coli* expression strains.

Transformation of competent bacteria with recombinant expression plasmids

Competent bacterial strains (BL21(DE3), BL21(DE3)pLyS, HMS174(DE3) and HMS174(DE3)pLysS were prepared and transformed with the recombinant pET28b expression plasmids carrying the cloned *H. pylori* sequences according to standard methods (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). These expression host strains contain a chromosomal copy of the gene for T7 RNA polymerase. These hosts are lysogens of bacteriophage DE3, a lambda derivate that carries the *lacI* gene, the *lacUV5* promoter and the gene for T7 RNA polymerase. T7 RNA polymerase expression is induced by the addition of isopropyl-β-D thiogalactoside (1PTG), and the T7 RNA polymerase then transcribes any target plasmid, such as pET28b, that carries a T7 promoter sequence and a gene of interest.

Expression of recombinant H. pylori sequences in E. coli

Transformants were collected from LB agar plates containing 25ug/ml kanamycin sulfate (ensures maintenance of the pET28b-based recombinant plasmids) and used to inoculate LB broth containing 25ug/ml kanamycin sulfate and grown to an optical density at 600nm of 0.5 to 1.0 OD units, at which point 1mM 1PTG was added to the culture for one to three hours to induce gene expression of the *H. pylori* recombinant DNA constructions. After induction of gene expression with 1PTG, bacteria were pelleted by centrifugation and resuspended in SDS-PAGE solubilization buffer and subjected to SDS-PAGE (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). Proteins were visualized by staining with Coomassie Brilliant Blue or detected by western immunoblotting using the specific anti-His tag monoclonal antibody (Clontech, Palo Alto, CA, USA) using standard methods (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., eds., 1994). The host strain that provided the highest level of recombinant protein production was then chosen for use in a large-scale induction in order to purify the recombinant protein. All of the following proteins listed were expressed recombinantly and the strain giving the highest level of expression listed: BL21(DE3) (vac31, vac26, vac37); BL21(DE3) pLysS (vac 9, 32); HMS174(DE3) (vac10,11).

Purification of recombinant proteins and generation of specific antiserum

Large scale cultures were inoculated and grown as above, and induced with 1mM IPTG for 3 hours. After induction, bacteria were pelleted by centrifugation in a Sorvall centrifuge at 3500 x g for 15 min at 4°C. All of the expressed recombinant proteins were present in the insoluble inclusion body fraction. Inclusion bodies were purified according to standard protocols (Antibodies, Cold Spring Harbor Laboratory Press, E. Harlow and D. Lane, eds., 1988). The recombinant protein produced by vac32 was solubilized in 8M urea and partially purified by nickel chromatography (REF here). Denatured recombinant proteins were purified by electrophoresis on SDS-PAGE gels, and after visualization with Coomassie Brilliant Blue, the protein was excised from the gel and the gel slices homogenized. This material was used to raise specific polyclonal antibodies in mice or rabbits according to standard protocols (Antibodies, Cold Spring Harbor Laboratory Press, E. Harlow and D. Lane, eds., 1988).

Immunological characterization of recombinant proteins

In all cases where antibody was attempted to be raised, high titre antisera was generated, confirming the immunogenicity of the recombinant proteins. Further, these specific antisera were used to analyze whether the protein encoded by the cloned gene was expressed in *H. pylori*. Western immunoblot analysis using standard protocols (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994) confirmed that the *H. pylori* strain J99 did express proteins of the expected molecular weight that reacted with the vac10, vac32, vac31, vac36 antiserum. The specific antiserum was also used to determine the level of antigenic conservation between a large number of *H. pylori* isolates that had been obtained from distinct geographical sites around the world, and from all types of clinical manifestations, including gastritis, duodenal ulcer, gastric ulcer and gastric cancer. It was found that every strain produced a protein that reacted specifically with each antiserum.

Further, *H. pylori* cells from strains J99, 17874, AH244 and SS1 were fractionated into different cellular compartments (Doig and Trust 1994 Infect. Immun. 62:4526-4533: O'Toole *et al.* 1995 J. Bacteriol. 177:6049-6057). The specific antiserum was used to probe .hese fractions by western immunoblot to identify in which fraction the protein was localized. In all cases, the immunoreactive protein was present in the outer membrane as had been predicted by the sequence features and motif searches described herein.

Demonstration of protein efficacy as a vaccine

Purification of vac36 for efficacy studies

All the following steps were carried out at 4°C. Cell pellets were resuspended in 5 volumes per gram of cell of lysis buffer (50mM Sodium Phosphate pH 8.0, 0.5 M NaCl, 5mM Imidazole) with 10mM EDTA, 1 mM phenylmethylsulfonyl fluoride (PMSF) and 0.1 % β-mercaptoethanol, and ruptured by several passages through a small volume microfluidizer (Model M-110S, Microfluidics International Corporation, Newton, MA). The resulting homogenate was made 0.2% sodium deoxycholate (DOC), stirred 20 minutes, then centrifuged (10,000 g x 30 min). The pellets were washed twice with Lysis Buffer containing 10mM EDTA, 1% Triton X-100, 1 mM PMSF and 0.1% β -mercaptoethanol, then with lysis buffer containing 1M urea, 1mM PMSF and 0.1 % β-mercaptoethanol. The resulting white pellet is composed primarily of inclusion bodies, free of unbroken cells and membranous materials.

The inclusion bodies were dissolved in 20 ml 6M guanidine-HCl in lysis buffer with 1 mM PMSF and 0.1% β-mercaptoethanol, and incubated on ice for 1 hour. Materials that did not dissolve were removed by centrifugation (100,000 g x 30 min.) The clear supernatant was filtered through a 0.8 µm Supor filter (Gelman Sciences, FRG) and then load directly onto a 10 ml Ni^{2+} - NTA agarose column (Hochuli et al. 1987) pre-equilibrated in 6M guanidine-HCl in Lysis Buffer containing 1 mM PMSF and 0.1% β-Mercaptoethanol. The column was washed with 20ml (2 bed volumes) of Lysis Buffer containing 6M guanidine-HCl, 1mM PMSF and 0.1 %β-mercaptoethanol, then guanidine-HCl was removed slowly with a 100 ml linear gradient (from 6M to 0 M Guanidine-HCl) of lysis buffer containing 0.5% Brij 35, 1 mM PMSF, 0.1% β-mercaptochanol. Next, the column was developed with a 25 ml linear gradient of increasing imidazole (5 to 500 mM) in Lysis buffer containing 0.5% Brij 35, 1 mM PMSF and 0.1% β-mercaptoethanol. The recombinant proteins elute as a peak centered at 100mM imidazole.

Fractions containing the recombinant proteins were pooled and then concentrated to approximately 8 ml by centrifugal filtration (Centriprep-10, Amicon, MA), and loaded directly onto a 350-ml column (2.2 X 91 cm) of Sephacyl S-100 HR gel filtration medium equilibrated in Buffer A (50mM Sodium Phosphate, pH 8.0, 500 mM NaCl, 0.1 mM EGTA, 1 mM PMSF, 0.1%β-mercaptoethanol, 0.5% Brij 35) and ran in Buffer A at 30 ml/h. Fractions containing the recombinant protein were identified by absorbance at 280 nm and analyzed by SDS-PAGE. Fractions were pooled, concentrated to 1.5 to 2 mg/ml and dialysed overnight against 10 mM Potassium Phosphate pH 7.5, 150 mM NaCl, 0.1 mM EGTA and 0.5% Brij 35. The concentration of protein in the dialysate was quantified, then aliquoted prior to freezing at - 20°C.

Mouse model of *Heliocobacter pylori* infection

A murine model of *H. pylori* infection was produced by infection of C57BL/6 mice with with *H. pylori* Sydney strain SS1 and was used to assess the efficacy of recombinant *H. pylori* vac36. This mouse-adapted *H. pylori* strain is *cagA+ vacA+*, shows colonization levels in C57BL/6 mice equivalent to those observed in humans, forms adhesion pedestals, colonizes for at least 8 months, and elicits a chronic-active gastritis and mucosal atrophy (Lee *et al.*, Gastroenterology, 112:1386-1397, 1997). Dose-response studies have shown 100% infection rates of inbred C57BL/6 and Balb/C mice at 8 weeks post-challenge with a single inoculation of 10^6 organisms.

Assessment of gastric *H. pylori* infection

The presence of *H. pylori* organisms in gastric tissue was determined by culture of gastric tissue and by a quantitative urease assay. In the latter method, a longitudinal segment of antrum, representing approximately ¼ of the total antral region was placed in 1 ml of urea broth. After 4 hr, the extent of color change resulting from urea hydrolysis and increased pH was quantiated by spectrophotometric measurement of A_{550} (Fox *et al.*, Immunol. 88:400-406, 1996). The assay sensitivity is ~ 10^3 *H. pylori* organisms. A positive (*H. pylori*-infected) gastric tissue was defined as that sample showing 2 standard deviations above the mean A_{550} value derived from a group of unchallenged uninfected age-matched control mice.

Assessment of local immune response to immunization in gastric tissue

Longitudinal sections of gastric tissues from the esophageal to the duodenal junction were embedded in OCT embedding compound, frozen in liquid nitrogen, and cryosections immunostained with monoclonal antibodies recognizing $CD4^+$ or $CD8^+$T cells or with antisera against mouse IgA for identification of IgA containing (IgACC) plasma cells (Pappo *et al.*, Infect. Immun. 63:1246-1252, 1995). The degree of local gastric immune response was expressed quantitatively as the number of $CD4^+$, $CD8^+$ or IgACC cells per mm^2 of gastric region examined.

Protective activity of purified recombinant *H. pylori* vac36 antigen

The ability of purified recombinant vac36 antigen derived from *H. pylori* to interfere with the establishment of an *H. pylori* infection was examined in mice. Groups (n=10) of 6-8 week-old female C57BL/6 mice were immunized orally 4 times at weekly intervals as follows: 1) 100 μg of recombinant vac36 antigen and 10 μg cholera toxin (CT) adjuvant, 2) 1 mg *H. pylori* lysate antigens and 10 μg CT, and 3) 0.2 M

bicarbonate buffer and 10 ug CT adjuvant. The mice were challenged 2 weeks later on 3 consecutive days by oral administration of 10^8 *H. pylori* organisms. The experiment was terminated 2 weeks post-challenge, and the *H. pylori* infection level assessed by bacterial colony counts and by quantitative urease assays.

Oral immunization with vac36 antigen interfered with the establishment of *H. pylori* infection upon challenge with live *H. pylori* organisms. Mice immunized with purified recombinant vac36 antigen exhibited a significantly lower level of colonization by *H. pylori*, as assessed by gastric urease activity and bacterial count assays (Table 6). Oral immunization with vac36 antigen also resulted in the generation of a local protective gastric immune response. Greater numbers of $CD4^+$T cells and of IgACC were recruited in the gastric tissues of vac36-immunized mice when compared with unimmunized *H. pylori*-infected mice (Table 7).

Table 6

Recombinant vac36 antigen protects mice from challenge with *H. pylori*

Vaccine Treatment Group	*Urease Activity*[a]	*p*[b]	*H. pylori burden*[c]	*p*[b]
vac36	0.199±0.080	0.0022	55,800±12,599	0.0125
H. pylori lysate	0.057±0.007	0.0002	2,360±955	0.0002
buffer	1.655±0.420	-	131,000±18,391	-

a Urease activity is expressed as mean A_{550}± SEM of duplicate antral samples from n=10 mice/group.

b by Wilcoxon Rank Sum Test compared with mice immunized with CT adjuvant alone

c The level of *H. pylori* in gastric tissue was assessed by bacterial counts, and shown as mean colony forming units±SEM

Table 7

vac36-immunized mice generate a local gastric immune response upon challenge with *H. pylori*

Vaccine Treatment Group	*CD4+*			*CD8+*			*IgACC*		
	cardia[a]	*corpus*	*antrum*	*cardia*	*corpus*	*antrum*	*cardia*	*corpus*	*antrum*
vac36	33 ± 9[a]	54 ± 8*	31 ± 8	3 ± 2	0	1 ± 1	24 ± 12	79 ± 16	67 ± 13
H. pylori lysate	31 ± 13	36 ± 19	24 ± 8	4 ± 2	2 ± 1	2 ± 1	31 ± 9	73 ± 13*	79 ± 15
buffer	12 ± 2	27 ± 8	18 ± 4	1 ± 1	0	0	4 ± 2	30 ± 13	46 ± 14

a Mean number of cells/mm^2 of gastric region ± SEM

* $p<0.05$ by Wilcoxon Rank Sum Test when compared with unimmunized *H. pylori* infected mice

V. Sequence Variance Analysis of genes in *Helicobacter pylori* strains

Four genes were cloned and sequenced from several strains of *H. pylori* to compare the DNA and deduced amino acid sequences. This information was used to determine the sequence variation between the *H. pylori* strain, J99, and other *H. pylori* strains isolated from human patients.

Preparation of Chromosomal DNA.

Cultures of *H. pylori* strains (as listed in Table 10) were grown in BLBB (1% Tryptone, 1% Peptamin 0.1% Glucose, 0.2% Yeast Extract 0.5% Sodium Chloride, 5% Fetal Bovine Serum) to an OD_{600} of 0.2. Cells were centrifuged in a Sorvall RC-3B at 3500 x g at 4^{o}C for 15 minutes and the pellet resuspended in 0.95 mls of 10 mM Tris-HCl, 0.1 mM EDTA (TE). Lysozyme was added to a final concentration of 1mg/ml along with, SDS to 1% and RNAse A + T1 to 0.5mg/ml and 5 units/ml respectively, and incubated at 37^{o}C for one hour. Proteinase K was then added to a final concentration of 0.4mg/ml and the sample was incubated at 55 C for more than one hour. NaCl was added to the sample to a concentration of 0.65 M, mixed carefully, and 0.15 ml of 10%

CTAB in 0.7M NaCL (final is 1% CTAB/70mM NaCL) was added followed by incubation at 65°C for 20 minutes. At this point, the samples were extracted with chloroform:isoamyl alcohol, extracted with phenol, and extracted again with chloroform:isoamyl alcohol. DNA was precipitated with either EtOH (1.5 x volumes) or isopropanol (0.6 x volumes) at -70°C for 10minutes, washed in 70% EtOH and resuspended in TE.

PCR Amplification and cloning.

Genomic DNA prepared from twelve strains of *Helicobacter pylori* was used as the source of template DNA for PCR amplification reactions (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994). To amplify a DNA sequence containing an *H. pylori* ORF, genomic DNA (10 nanograms) was introduced into a reaction vial containing 2 mM $MgCl_2$, 1 micromolar synthetic oligonucleotide primers (forward and reverse primers, see Table 8) complementary to and flanking a defined *H. pylori* ORF, 0.2 mM of each deoxynucleotide triphosphate; dATP, dGTP, dCTP, dTTP and 0.5 units of heat stable DNA polymerase (Amplitaq, Roche Molecular Systems, Inc., Branchburg, NJ, USA) in a final volume of 20 microliters in duplicate reactions.

Table 8
Oligonucleotide primers used for PCR amplification of *H. pylori* DNA sequences.

Outer membrane Proteins	Forward primer 5' to 3'	Reverse Primer 5' to 3'
Protein 26054702 (for strains AH4, AH15, AH61, 5294, 5640, AH18, and AH244)	5'-TTAACCATGGTGAAAA GCGATA-3' (SEQ ID NO:217)	5'-TAGAATTCGCCTCTAA AACTTTAG-3' (SEQ ID NO:218)
Protein 26054702 (for strains AH5, 5155, 7958, AH24,and J99)	5'-TTAACCATGGTGAAAA GCGATA-3' (SEQ ID NO:219)	5'-TAGAATTCGCATAA CGATCAATC-3' (SEQ ID NO:220)
Protein 7116626	5'-ATATCCATGGTGAGTT TGATGA-3' (SEQ ID NO:221)	5'-ATGAATTCAATTTT TTATTTTGCCA-3' (SEQ ID NO:222)
Protein 29479681	5'-AATTCCATGGCTATC CAAATCCG-3' (SEQ ID NO:223)	5'-ATGAATTCGCCAAAA TCGTAGTATT-3' (SEQ ID NO:224)
Protein 346	5'-GATACCATGGAATTT ATGAAAAAG-3' (SEQ ID NO:225)	5'-TGAATTCGAAAAAGTG TAGTTATAC-3' (SEQ ID NO:226)

The following thermal cycling conditions were used to obtain amplified DNA products for each ORF using a Perkin Elmer Cetus/ GeneAmp PCR System 9600 thermal cycler:

Protein 7116626 and Protein 346;
Denaturation at 94°C for 2 min,
2 cycles at 94°C for 15 sec, 30°C for 15 sec and 72°C for 1.5 min
23 cycles at 94°C for 15 sec, 55°C for 15 sec and 72°C for 1.5 min
Reactions were concluded at 72°C for 6 minutes.

Protein 26054702 for strains AH5, 5155, 7958, AH24,and J99;
Denaturation at 94°C for 2 min,
2 cycles at 94°C for 15 sec, 30°C for 15 sec and 72°C for 1.5 min
25 cycles at 94°C for 15 sec, 55°C for 15 sec and 72°C for 1.5 min
Reaction was concluded at 72°C for 6 minutes.

Protein 26054702 and Protein 294796813 for strains AH4, AH15, AH61, 5294, 5640, AH18, and Hp244 ;
Denaturation at 94°C for 2 min,
2 cycles at 94°C for 15 sec, 30°C for 20 sec and 72°C for 2 min
25 cycles at 94°C for 15 sec, 55°C for 20 sec and 72°C for 2 min
Reactions were concluded at 72°C for 8 minutes.

Upon completion of thermal cycling reactions, each pair of samples were combined and used directly for cloning into the pCR cloning vector as described below.

Cloning of H. pylori DNA sequences into the pCR TA cloning vector.

All amplified inserts were cloned into the pCR 2.1 vector by the method described in the Original TA cloning kit (Invitrogen, San Diego, CA). Products of the ligation reaction were then used to transform the TOP10F' (INVaF' in the case of *H. pylori* sequence 350) strain of *E. coli* as described below.

Transformation of competent bacteria with recombinant plasmids

Competent bacteria, *E coli* strain TOP10F' or *E. coli* strain INVaF' were transformed with recombinant pCR expression plasmids carrying the cloned *H. pylori* sequences according to standard methods (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994). Briefly, 2 microliters of 0.5

micromolar BME was added to each vial of 50 microliters of competent cells. Subsequently, 2 microliters of ligation reaction was mixed with the competent cells and incubated on ice for 30 minutes. The cells and ligation mixture were then subjected to a "heat shock" at 42°C for 30 seconds, and were subsequently placed on ice for an additional 2 minutes, after which, samples were incubated in 0.45 milliliters SOC medium (0.5% yeast extract, 2.0 % tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl2, 10 mM MgSO4 and 20, mM glucose) at 37°C with shaking for 1 hour. Samples were then spread on LB agar plates containing 25 microgram/ml kanamycin sulfate or 100 micrograms/ml ampicillan for growth overnight. Transformed colonies of TOP10F' or INVaF' were then picked and analyzed to evaluate cloned inserts as described below.

Identification of recombinant PCR plasmids carrying H. pylori sequences

Individual TOP10F' or INVaF' clones transformed with recombinant pCR-*H.pylori* ORFs were analyzed by PCR amplification of the cloned inserts using the same forward and reverse primers, specific for each *H. pylori* sequence, that were used in the original PCR amplification cloning reactions. Successful amplification verified the integration of the *H. pylori* sequences in the cloning vector (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994).

Individual clones of recombinant pCR vectors carrying properly cloned *H. pylori* ORFs were picked for sequence analysis. Sequence analysis was performed on ABI Sequencers using standard protocols (Perkin Elmer) using vector-specific primers (as found in PCRII or pCR2.1, Invitrogen, San Diego, CA) and sequencing primers specific to the ORF as listed in Table 9 below.

Table 9
Oligonucleotide primers used for sequencing of *H. pylori* DNA sequences.

Outer membrane Proteins	Forward primers 5' to 3'	Reverse Primers 5' to 3'
Protein 26054702	5'CCCTTCATTTTAGAAATCG-3' (SEQ ID NO:227) 5'ATTTCAACCAATTCAATGCG-3' (SEQ ID NO:228) 5'GCCCCTTTTGATTTGAAGCT-3' (SEQ ID NO:229) 5'TCGCTCCAAGATACCAAGAAGT-3' (SEQ ID NO:230) 5'CTTGAATTAGGGGCAAAGATCG-3' (SEQ ID NO:231) 5'ATGCGTTTTTACCCAAAGAAGT-3' (SEQ ID NO:232) 5'ATAACGCCACTTCCTTATTGGT-3' (SEQ ID NO:233)	5'CTTTGGGTAAAAACGCATC-3' (SEQ ID NO:234) 5'CGATCTTTGATCCTAATTCA-3' (SEQ ID NO:235) 5'ATCAAGTTGCCTATGCTGA-3' (SEQ ID NO:236)
Protein 7116626	5'TTGAACACTTTTGATTATGCGG-3' (SEQ ID NO:237) 5'GGATTATGCGATTGTTTTACAAG-3' (SEQ ID NO:238)	5'GTCTTTAGCAAAAATGGCGTC-3' (SEQ ID NO:239) 5'AATGAGCGTAAGAGAGCCTTC-3' (SEQ ID NO:240)
Protein 29479681	5'CTTATGGGGGTATTGTCA-3' (SEQ ID NO:241) 5'AGCATGTGGGTATCCAGC-3' (SEQ ID NO:242)	5'AGGTTGTTGCCTAAAGACT-3' (SEQ ID NO:243) 5'-CTGCCTCCACCTTTGATC-3' (SEQ ID NO:244)
Protein 346	5'ACCAATATCAATTGGCACT-3' (SEQ ID NO:245) 5'ACTTGGAAAAGCTCTGCA-3' (SEQ ID NO:246)	5'CTTGCTTGTCATATCTAGC-3' (SEQ ID NO:247) 5'-GTTGAAGTGTTGGTGCTA-3' (SEQ ID NO:248)
	5'CAAGCAAGTGGTTTGGTTTTAG-3' (SEQ ID NO:249) 5'TGGAAAGAGCAAATCATTGAAG-3' (SEQ ID NO:250)	5'GCCCATAATCAAAAAGCCCAT-3' (SEQ ID NO:251) 5'CTAAAACCAAACCACTTGCTTGTC-3' (SEQ ID NO:252)
Vector Primers	5'-GTAAAACGACGGCCAG-3' (SEQ ID NO:253)	5'-CAGGAAACAGCTATGAC-3' (SEQ ID NO:254)

Results

To establish the PCR error rate in these experiments, five individual clones of Protein 26054702, prepared from five separate PCR reaction mixtures from *H. pylori* strain J99, were sequenced over a total length of 897 nucleotides for a cumulative total of 4485 bases of DNA sequence. DNA sequence for the five clones was compared to a DNA sequence obtained previously by a different method, i.e., random shotgun cloning and sequencing. The PCR error rate for the experiments described herein was determined to be 2 base changes out of 4485 bases, which is equivalent to an estimated error rate of less than or equal to 0.04%.

DNA sequence analysis was performed on four different open reading frames identified as genes and amplified by PCR methods from a dozen different strains of the bacterium *Helicobacter pylori*. The deduced amino acid sequences of three of the four open reading frames that were selected for this study showed statistically significant BLAST homology to defined proteins present in other bacterial species. Those ORFs included: Protein 26054702, homologous to the val A & B genes encoding an ABC transporter in F. novicida; Protein 7116626, homologous to lipoprotein e (P4) present in the outer membrane of H. influenzae; Protein 29479681, homologous to fecA, an outer membrane receptor in iron (III) dicitrate transport in *E. coli*. Protein 346 was identified as an unknown open reading frame, because it showed low homology with sequences in the public databases.

To assess the extent of conservation or variance in the ORFs across various strains of *H. pylori*, changes in DNA sequence and the deduced protein sequence were compared to the DNA and deduced protein sequences found in the J99 strain of *H. pylori* (*see* Table 10 below). Results are presented as percent identity to the J99 strain of *H. pylori* sequenced by random shotgun cloning. To control for any variations in the J99 sequence each of the four open reading frames were cloned and sequenced again from the J99 bacterial strain and that sequence information was compared to the sequence information that had been collected from inserts cloned by random shotgun sequencing of the J99 strain. The data demonstrate that there is variation in the DNA sequence ranging from as little as 0.12 % difference (Protein 346, J99 strain) to approximately 7% change (Protein 26054702, strain AH5). The deduced protein sequences show either no variation (Protein 346, strains AH18 and AH24) or up to as much as 7.66% amino acid changes (Protein 26054702, Strain AH5).

Table 10

Multiple Strain DNA Sequence analysis of H. pylori Vaccine Candidates

J99 Protein #:	26054702	26054702	7116626	7116626	29479681	29479681	346	346
Length of Region Sequenced:	248 a.a.	746 nt.	232 a.a.	96 nt.	182 a.a.	548 nt.	273 a.a.	819 nt.
Strain Tested	AA identity	Nuc. identity	AA identity	Nuc. identity	AA identity	Nuc. identity	AA identity	Nuc. identity
J99	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	99.63%	99.88%
AH244	95.16%	95.04%	n.d.	n.d.	99.09%	96.71%	98.90%	96.45%
AH4	95.97%	95.98%	97.84%	95.83%	n.d.	n.d.	97.80%	95.73%
AH5	92.34%	93.03%	98.28%	96.12%	98.91%	96.90%	98.53%	95.73%
AH15	95.16%	94.91%	97.41%	95.98%	99.82%	97.99%	99.63%	96.09%
AH61	n.d.	n.d.	97.84%	95.98%	99.27%	97.44%	n.d.	n.d.
5155	n.d.	n.d.	n.d.	n.d.	99.45%	97.08%	98.53%	95.60%
5294	94.35%	94.37%	98.28%	95.40%	99.64%	97.26%	97.07%	95.48%
7958	94.35%	94.10%	97.84%	95.40%	n.d.	n.d.	99.63%	96.46%
5640	95.16%	94.37%	97.41%	95.69%	99.09%	97.63%	98.53%	95.48%
AH18	n.d.	n.d.	98.71%	95.69%	99.64%	97.44%	100.00%	95.97%
AH24	94.75%	95.04%	97.84%	95.40%	99.27%	96.71%	100.00%	96.46%

n.d.= not done.

VI. Experimental Knock-Out Protocol for the Determination of Essential *H. pylori* Genes as Potential Therapeutic Targets

Therapeutic targets are chosen from genes whose protein products appear to play key roles in essential cell pathways such as cell envelope synthesis, DNA synthesis, transcription, translation, regulation and colonization/virulence.

The protocol for the deletion of portions of *H. pylori* genes/ORFs and the insertional mutagenesis of a kanamycin-resistance cassette in order to identify genes which are essential to the cell is modified from previously published methods (Labigne-Roussel et al., 1988, J. Bacteriology 170, pp. 1704-1708; Cover et al.,1994, J. Biological

Chemistry 269, pp. 10566-10573; Reyrat et al., 1995, Proc. Natl. Acad. Sci. 92, pp 8768-8772). The result is a gene "knock-out."

Identification and Cloning of H. pylori Gene Sequences

The sequences of the genes or ORFs (open reading frames) selected as knock-out targets are identified from the *H. pylori* genomic sequence and used to design primers to specifically amplify the genes/ORFs. All synthetic oligonucleotide primers are designed with the aid of the OLIGO program (National Biosciences, Inc., Plymouth, MN 55447, USA), and can be purchased from Gibco/BRL Life Technologies (Gaithersburg, MD, USA). If the ORF is smaller than 800 to 1000 base pairs, flanking primers are chosen outside of the open reading frame.

Genomic DNA prepared from the *Helicobacter pylori* HpJ99 strain (ATCC 55679; deposited by Genome Therapeutics Corporation, 100 Beaver Street, Waltham, MA 02154) is used as the source of template DNA for amplification of the ORFs by PCR (polymerase chain reaction) (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994). For the preparation of genomic DNA from *H. pylori*, see Example I. PCR amplification is carried out by introducing 10 nanograms of genomic HpJ99 DNA into a reaction vial containing 10 mM Tris pH 8.3, 50 mM KCl, 2 mM $MgCl_2$, 2 microMolar synthetic oligonucleotide primers (forward=F1 and reverse=R1), 0.2 mM of each deoxynucleotide triphosphate (dATP,dGTP, dCTP, dTTP), and 1.25 units of heat stable DNA polymerase (Amplitaq, Roche Molecular Systems, Inc., Branchburg, NJ, USA) in a final volume of 40 microliters. The PCR is carried out with Perkin Elmer Cetus/GeneAmp PCR System 9600 thermal cyclers.

Upon completion of thermal cycling reactions, each sample of amplified DNA is visualized on a 2% TAE agarose gel stained with Ethidium Bromide (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994) to determine that a single product of the expected size had resulted from the reaction. Amplified DNA is then washed and purified using the Qiaquick Spin PCR purification kit (Qiagen, Gaithersburg, MD, USA).

PCR products a.e cloned into the pT7Blue T-Vector (catalog#69820-1, Novagen, Inc., Madison, WI, USA) using the TA cloning strategy (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994). The ligation of the PCR product into the vector is accomplished by mixing a 6 fold molar excess of the PCR product, 10 ng of pT7Blue-T vector (Novagen), 1 microliter of T4 DNA Ligase Buffer (New England Biolabs, Beverly, MA, USA), and 200 units of T4 DNA Ligase

(New England Biolabs) into a final reaction volume of 10 microliters. Ligation is allowed to proceed for 16 hours at 16°C.

Ligation products are electroporated (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994) into electroporation-competent XL-1 Blue or DH5-a *E.coli* cells (Clontech Lab., Inc. Palo Alto, CA, USA). Briefly, 1 microliter of ligation reaction is mixed with 40 microliters of electrocompetent cells and subjected to a high voltage pulse (25 microFarads, 2.5 kV, 200 ohms) after which the samples are incubated in 0.45 ml SOC medium (0.5% yeast extract, 2% tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$ and 20 mM glucose) at 37°C with shaking for 1 hour. Samples are then spread onto LB (10 g/l bacto tryptone, 5 g/l bacto yeast extract, 10 g/l sodium chloride) plates containing 100 microgram/ml of Ampicillin, 0.3% X-gal, and 100 microgram/ml IPTG. These plates are incubated overnight at 37°C. Ampicillin-resistant colonies with white color are selected, grown in 5 ml of liquid LB containing 100 microgram/ml of Ampicillin, and plasmid DNA is isolated using the Qiagen miniprep protocol (Qiagen, Gaithersburg, MD, USA).

To verify that the correct *H.pylori* DNA inserts had been cloned, these pT7Blue plasmid DNAs are used as templates for PCR amplification of the cloned inserts, using the same forward and reverse primers used for the initial amplification of the J99 *H.pylori* sequence. Recognition of the primers and a PCR product of the correct size as visualized on a 2% TAE, ethidium bromide stained agarose gel are confirmation that the correct inserts had been cloned. Two to six such verified clones are obtained for each knock-out target, and frozen at -70°C for storage. To minimize errors due to PCR, plasmid DNA from these verified clones are pooled, and used in subsequent cloning steps.

The sequences of the genes/ORFs are again used to design a second pair of primers which flank the region of *H. pylori* DNA to be either interrupted or deleted (up to 250 basepairs) within the ORFs but are oriented away from each other. The pool of circular plasmid DNAs of the previously isolated clones are used as templates for this round of PCR. Since the orientation of amplification of this pair of deletion primers is away from each other, the portion of the ORF between the primers is not included in the resultant PCR product. The PCR product is a linear piece of DNA with *H. pylori* DNA at each end and the pT7Blue vector backbone between them which, in essence, resultes in the deletion of a portion of the ORFs. The PCR product is visualized on a 1% TAE, ethidium bromide stained agarose gel to confirm that only a single product of the correct size has been amplified.

A Kanamycin-resistance cassette (Labigne-Roussel et al., 1988 J. Bacteriology 170, 1704-1708) is ligated to this PCR product by the TA cloning method used previously (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel et al., editors, 1994). The Kanamycin cassette containing a *Campylobacter* kanamycin resistance gene is obtained by carrying out an EcoRI digestion of the recombinant plasmid pCTB8:*kan* (Cover et al.,1994, J. Biological Chemistry 269, pp. 10566-10573). The proper fragment (1.4 kb) is isolated on a 1% TAE gel, and isolated using the QIAquick gel extraction kit (Qiagen, Gaithersburg, MD, USA). The fragment is end repaired using the Klenow fill-in protocol, which involved mixing 4ug of the DNA fragment, 1 microliter of dATP,dGTP, dCTP, dTTP at 0.5 mM, 2 microliter of Klenow Buffer (New England Biolabs) and 5 units of Klenow DNA Polymerase I Large (Klenow) Fragment (New England Biolabs) into a 20 microliter reaction, incubating at 30°C for 15 min, and inactivating the enzyme by heating to 75°C for 10 minutes. This blunt-ended Kanamycin cassette is then purified through a Qiaquick column (Qiagen, Gaithersburg, MD, USA) to eliminate nucleotides. The "T" overhang is then generated by mixing 5 micrograms of the blunt-ended kanamycin cassette, 10 mM Tris pH 8.3, 50 mM KCl, 2 mM $MgCl_2$, 5 units of DNA Polymerase (Amplitaq, Roche Molecular Systems, Inc., Branchburg, NJ, USA), 20 microliters of 5 mM dTTP, in a 100 microliter reaction and incubating the reaction for 2 hours at 37°C. The "Kan-T" cassette is purified using a QIAquick column (Qiagen, Gaithersburg, MD, USA). The PCR product of the deletion primers (F2 and R2) is ligated to the Kan-T cassette by mixing 10 to 25 ng of deletion primer PCR product, 50 - 75 ng Kan-T cassette DNA, 1 microliter 10x T4 DNA Ligase reaction mixture, 0.5 microliter T4 DNA Ligase (New England Biolabs, Beverly, MA, USA) in a 10 microliter reaction and incubating for 16 hours at 16°C.

The ligation products are transformed into XL-1 Blue or DH5-a *E.coli* cells by electroporation as described previously. After recovery in SOC, cells are plated onto LB plates containing 100 microgram/ml Ampicillin and grown overnight at 37°C. These plates are then replica plated onto plates containing 25 microgram/ml Kanamycin and allowed to grow overnight. Resultant colonies have both the Ampicillin resistance gene present in the pT7Blue vector, and the newly introduced Kanamycin resistance gene. Colonies are picked into LB containing 25 microgram/ml Kanamycin and plasmid DNA is isolated from the cultured cells using the Qiagen miniprep protocol (Qiagen, Gaithersburg, MD, USA).

Several tests by PCR amplification are conducted on these plasmids to verify that the Kanamycin is inserted in the *H. pylori* gene/ORF, and to determine the orientation of the insertion of the Kanamycin-resistance gene relative to the *H. pylori* gene/ORF. To

verify that the Kanamycin cassette is inserted into the *H. pylori* sequence, the plasmid DNAs are used as templates for PCR amplification with the set of primers originally used to clone the *H. pylori* gene/ORFs. The correct PCR product is the size of the deleted gene/ORF but increased in size by the addition of a 1.4 kilobase Kanamycin cassette. To avoid potential polar effects of the kanamycin resistance cassette on *H. pylori* gene expression, the orientation of the Kanamycin resistance gene with respect to the knock-out gene/ORF is determined and both orientations are eventually used in *H. pylori* transformations (see below). To determine the orientation of insertion of the kanamycin resistance gene, primers are designed from the ends of the kanamycin resistance gene ("Kan-1" 5'-ATCTTACCTATCACCTCAAAT-3' (SEQ ID NO:255)), and "Kan-2" 5'-AGACAGCAACATCTTTGTGAA-3' (SEQ ID NO:256)). By using each of the cloning primers in conjunction with each of the Kan primers (4 combinations of primers), the orientation of the Kanamycin cassette relative to the *H.pylori* sequence is determined. Positive clones are classified as either in the "A" orientation (the same direction of transcription is present for both the *H. pylori* gene and the Kanamycin resistance gene), or in the "B" orientation (the direction of transcription for the *H.pylori* gene is opposite to that of the Kanamycin resistance gene). Clones which share the same orientation (A or B) are pooled for subsequent experiments and independently transformed into *H. pylori*.

Transformation of Plasmid DNA into H. pylori cells

Two strains of *H. pylori* are used for transformation: ATCC 55679, the clinical isolate which provided the DNA from which the *H. pylori* sequence database is obtained, and AH244, an isolate which had been passaged in, and has the ability to colonize the mouse stomach. Cells for transformation are grown at 37°C, 10% CO_2, 100% humidity, either on Sheep-Blood agar plates or in Brucella Broth liquid. Cells are grown to exponential phase, and examined microscopically to determine that the cells are "healthy" (actively moving cells) and not contaminated. If grown on plates, cells are harvested by scraping cells from the plate with a sterile loop, suspended in 1 ml of Brucella Broth, spun down (1 minute, top speed in eppendorf microfuge) and resuspended in 200 microliters Brucella Broth. If grown in Brucella Broth liquid, cells are centrifuged (15 minutes at 3000 rpm in a Beckman TJ6 centrifuge) and the cell pellet resuspended in 200 microliters of Brucella broth. An aliquot of cells is taken to determine the optical density at 600 nm, in order to calculate the concentration of cells. An aliquot (1 to 5 OD_{600} units/25 microliter) of the resuspended cells is placed onto a prewarmed Sheep-Blood agar plate, and the plate is further incubated at 37°C, 6% CO_2,

100% humidity for 4 hours. After this incubation, 10 microliters of plasmid DNA (100 micrograms per microliter) is spotted onto these cells. A positive control (plasmid DNA with the ribonuclease H gene disrupted by kanamycin resistance gene) and a negative control (no plasmid DNA) are done in parallel. The plates are returned to 37°C, 6% CO_2 for an additional 4 hours of incubation. Cells are then spread onto that plate using a swab wetted in Brucella broth, and grown for 20 hours at 37°C, 6% CO_2. Cells are then transferred to a Sheep-Blood agar plate containing 25 micrograms/ml Kanamycin, and allowed to grow for 3 to 5 days at 37°C, 6% CO_2, 100% humidity. If colonies appear, they are picked and regrown as patches on a fresh Sheep-Blood agar plate containing 25 micrograms/ml Kanamycin.

Three sets of PCR tests are done to verify that the colonies of transformants have arisen from homologous recombination at the proper chromosomal location. The template for PCR (DNA from the colony) is obtained by a rapid boiling DNA preparation method as follows. An aliquot of the colony (stab of the colony with a toothpick) is introduced into 100 microliters of 1% Triton X-100, 20 mM Tris, pH 8.5, and boiled for 6 minutes. An equal volume of phenol : chloroform (1:1) is added and vortexed. The mixture is microfuged for 5 minutes and the supernatant is used as DNA template for PCR with combinations of the following primers to verify homologous recombination at the proper chromosomal location.

TEST 1. PCR with cloning primers originally used to amplify the gene/ORF. A positive result of homologous recombination at the correct chromosomal location should show a single PCR product whose size is expected to be the size of the deleted gene/ORF but increased in size by the addition of a 1.4 kilobase Kanamycin cassette. A PCR product of just the size of the gene/ORF is proof that the gene had not been knocked out and that the transformant is not the result of homologous recombination at the correct chromosome location.

TEST 2. PCR with F3 (primer designed from sequences upstream of the gene/ORF and not present on the plasmid), and either primer Kan-1 or Kan-2 (primers designed from the ends of the kanamycin resistance gene), depending on whether the plasmid DNA used was of "A" or "B" orientation. Homologous recombination at the correct chromosomal location will result in a single PCR product of the expected size (i.e., from the location of F3 to the insertion site of kanamycin resistance gene). No PCR product or PCR product(s) of incorrect size(s) will prove that the plasmid had not integrated at the correct site and that the gene had not been knocked out.

TEST 3. PCR with R3 (primer designed from sequences downstream of the gene/ORF and not present on the plasmid) and either primer Kan-1 or Kan-2, depending on whether the plasmid DNA used was of "A" or "B" orientation. Homologous

recombination at the correct chromosomal location will result in a single PCR product of the expected size (i.e., from the insertion site of kanamycin resistance gene to the downstream location of R3). Again, no PCR product or PCR product(s) of incorrect size(s) will prove that the plasmid had not integrated at the correct site and that the gene had not been knocked out.

Transformants showing positive results for all three tests above indicate that the gene is not essential for survival *in vitro*.

A negative result in any of the three above tests for each transformant indicates that the gene had not been disrupted, and that the gene is essential for survival *in vitro*.

In the event that no colonies result from two independent transformations while the positive control with the disrupted ribonuclease H plasmid DNA produces transformants, the plasmid DNA is further analyzed by PCR on DNA from transformant populations prior to plating for colony formation. This will verify that the plasmid can enter the cells and undergo homologous recombination at the correct site. Briefly, plasmid DNA is incubated according to the transformation protocol described above. DNA is extracted from the *H. pylori* cells immediately after incubation with the plasmid DNAs and the DNA is used as template for the above TEST 2 and TEST 3. Positive results in TEST 2 and TEST 3 would verify that the plasmid DNA could enter the cells and undergo homologous recombination at the correct chromosomal location. If TEST 2 and TEST 3 are positive, then failure to obtain viable transformants indicates that the gene is essential, and cells suffering a disruption in that gene are incapable of colony formation.

VII. High-throughput drug screen assay

Cloning, expression and protein purification

Cloning, transformation, expression and purification of the *H. pylori* target gene and its protein product,e.g., an *H. pylori* enzyme, to be used in a high-throughput drug screen assay, is carried out essentially as described in Examples II and III above. Development and application of a screening assay for a particular *H. pylori* gene product, peptidyl-propyl *cis-trans* isomerase, is described below as a specific example.

Enzymatic Assay

The assay is essentially as described by Fisher (Fischer, G., et.al. (1984) *Biomed. Biochim. Acta* 43:1101-1111). The assay measures the *cis-trans* isomerization of the Ala-Pro bond in the test peptide N-succinyl-Ala-Ala-Pro-Phe-p-nitroanilide (Sigma # S-7388, lot # 84H5805). The assay is coupled with α-chymotrypsin, where the ability of the protease to cleave the test peptide occurs only when the Ala-Pro bond is in *trans*. The conversion of the test peptide to the trans isomer in the assay is followed at 390 nm

on a Beckman Model DU-650 spectophotometer. The data are collected every second with an average scanning of time of 0.5 second. Assays are carried out in 35 mM Hepes, pH 8.0, in a final volume of 400 ul, with 10 μM α-chymotrypsin (type 1-5 from bovine Pancreas, Sigma # C-7762, lot 23H7020) and 10 nM PPIase. To initiate the reaction, 10 μl of the substrate (2 mM N-Succinyl-Ala-Ala-Pro-Phe-p-nitroanilide in DMSO) is added to 390 μl of reaction mixture at room temperature.

Enzymatic assay in crude bacterial extract.

A 50 ml culture of *Helicobacter pylori* (strain J99) in Brucella broth is harvested at mid-log phase ($OD_{600\ nm}$ ~ 1) and resuspended in lysis buffer with the following protease inhibitors: 1 mM PMSF, and 10 μg/ml of each of aprotinin, leupeptin, pepstatine, TLCK, TPCK, and soybean trypsin inhibitor. The suspension is subjected to 3 cycles of freeze-thaw (15 minutes at -70 ° C, then 30 minutes at room temperature), followed by sonication (three 20 second bursts). The lysate is centrifuged (12,000 g x 30 minutes) and the supernatant is assayed for enzymatic activity as described above.

Many *H. pylori* enzymes can be expressed at high levels and in an active form in *E. coli*. Such high yields of purified proteins provide for the design of various high throughput drug screening assays.

VIII. Truncated gene expression and protein production

Identification, cloning and expression of recombinant Helicobacter pylori sequences.

To facilitate the cloning, expression and purification of membrane proteins from *H. pylori*, the pET gene expression system (Novagen), for cloning and expression of recombinant proteins in *Escherichia coli* was selected. Further, for proteins that have a signal sequence at their amino-terminal end, a DNA sequence encoding a peptide tag (His-tag) was fused to the 5' end of the *H. pylori* DNA sequences of interest in order to facilitate purification of the recombinant protein products. In some cases, the DNA sequence was cloned in frame with the glutathione-S-transferase protein to produce a GST-fusion protein. The vectors used in this case were the pGEX series from Pharmacia LKB (Uppsala, Sweden).

PCR amplification and cloning of DNA sequences containing ORFs for membrane and secreted proteins from the J99 strain of Helicobacter pylori.

The sequences chosen (from the list of the DNA sequences of the invention) for cloning from *H. pylori* strain J99 were prepared for amplification cloning by the polymerase chain reaction (PCR). Synthetic oligonucleotide primers for the ORF of interest (Table 11) specific for the predicted mature 5' end of the ORF and either

downstream (3') of the predicted translational termination codon or at specific points within the coding region were designed and purchased (GibcoBRL Life Technologies, Gaithersburg, MD, USA). All forward primers (specific for the 5' terminus of the region of ORF of interest) were designed to include either a *Bam*HI or a *Nde*I restriction site. These primers within the *Nde*I restriction site sequence were designed to permit the initiation of protein translation at a methionine residue (encoded within the *Nde*I restriction site sequence, in the case of producing a non His-tagged recombinant protein) or to fuse in frame with the DNA sequence encoding the His-tag (for producing His tagged recombinant protein), followed by the coding sequence for the remainder of the native *H. pylori* DNA. The primer with the *Bam*HI restriction site was produced to fuse the *H. pylori* specific sequence in-frame with the C-terminus of the glutathione-S-transferase gene in the pGEX vectors (Pharmacia LKB, Uppsala, Sweden). All reverse oligonucleotide primers designed to include an *Eco*RI restriction site at the 5' terminus. Several reverse oligonucleotide primers were selected that would cause a truncation of the polypeptide to remove certain portions of the C-terminus, and in these cases the *Eco*RI restriction site at the 5' end was followed by a translational termination codon. This combination of primers would enable the ORF of interest (or parts of the ORF of interest) to be cloned into pET28b (to produce a His-tagged recombinant protein), pET30a (to produce a non His tagged or native recombinant protein) or the pGEX-4T or pGEX-5X series (to produce a GST fusion protein). The pET28b vector provides sequence encoding an additional 20 amino-terminal amino acids (plus the methionine in the *Nde*I restriction site) including a stretch of six histidine residues which makes up the His-tag, whereas the pGEX vectors fuse the *H. pylori* protein to a 26,000Da glutathione-S-transferase protein.

Genomic DNA prepared from *H. pylori* strain J99 (ATCC 55679) was used as the source of template DNA for the PCR amplification reactions (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). To amplify a DNA sequence containing a specific *H. pylori* ORF, genomic DNA (50 nanograms) was introduced into a reaction tube containing 200 nanograms of both the forward and reverse synthetic oligonucleotide primer specific for the ORF of interest, and 45 microliters of PCR SuperMix purchased (GibcoBRL Life Technologies, Gaithersburg, MD, USA) in a total of 50 microliters. The PCR SuperMix is supplied in 1.1X concentrations and contains 22mM Tris-HCl (pH 8.4), 55mM KCl, 1.65mM $MgCl_2$, 220 micromolar of each dATP, dCTP, dGTP and dTTP, 22units recombinant *Taq* polymerase/ml and stabilizers. The following thermal cycling conditions were used to obtain amplified DNA products for each ORF using a Perkin Elmer Cetus/GeneAmp PCR System thermal cycler.

Table 11: Oligonucleotide primers

Gene and location	Sequence
Vac38- BamHI post signal sequence	CGGGATCCGAAGGTGATGGTGTTTATA TAGG (SEQ ID NO: 271)
Vac38- NdeI post signal sequence	CGCATATGGAAGGTGATGGTGTTTATA TAGGG (SEQ ID NO: 272)
Vac38- EcoRI/stop codon (removes C-terminal third of protein)	GCGAATTCTCACTCTTTCCAATAGTTTG CTGCAGAGC (SEQ ID NO: 273)
Vac38- EcoRI/stop codon (removes C-terminal 11 amino acids)	CCGGAATTCTTAATCCCGTTTCAAATG GTAATAAAGG (SEQ ID NO: 274)
Vac38- EcoRI downstream of native stop codon	GCGAATTCCCTTTTATTTAAAAAGTGT AGTTATACC (SEQ ID NO: 275)

Sequences for Vac38 (full length or truncated)
Denaturation at 94°C for 30 sec
35 cycles at 94°C for 15 sec, 55°C for 15 sec, and 72°C for 1.5 min
Reactions were concluded at 72°C for 8 minutes

Upon completion of the thermal cycling reactions, each sample of amplified DNA was subjected to electrophoresis on 1.0% agarose gels. The DNA was visualized by exposure to ethidium bromide and long wave UV irradiation, and cut out in gel slices. DNA was purified using the Wizard PCR Preps kit (Promega Corp., Madison WI, USA), and then subjected to digestion with *Bam*HI and *Eco*RI (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). The digested PCR amplicon was then re-electrophoresed and purified as before.

Ligation of H. pylori DNA sequences into cloning vectors

The pOK12 vector (J. Vieira and J. Messing, Gene 100:189-194, 1991) was prepared for cloning by digestion with *Bam*HI and *Eco*RI or *Nde*I and *Eco*RI in the case of Vac41 (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). The vectors were subjected to electrophoresis on 1.0% agarose gels and purified using the Wizard PCR Preps kit (Promega Corp., Madison WI, USA). Following ligation of the purified, digested vector and the purified, digested

amplified *H. pylori* ORF, the products of the ligation reaction were transformed into *E. coli* JM109 competent cells according to standard methods (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). Individual bacterial colonies were screened for those containing the correct recombinant plasmids by incubating in LB broth overnight (plus 25ug/ml kanamycin sulfate) followed by plasmid DNA preparation using the Magic Minipreps system (Promega Corp., Madison WI, USA), and then analyzed by restriction digestion (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994).

Cloning of H. pylori DNA sequences into the pET28b, pET30a and pGEX4T-3 prokaryotic expression vectors

Both the pET28b and pET30a expression vectors were prepared for cloning by digestion with *Nde*I and *Eco*RI, and the pGEX4T-3 vector was prepared for cloning by digestion with *Bam*HI and *Eco*RI (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). The *H. pylori* DNA sequences were removed from pOK12 plasmid backbones by digestion with *Nde*I and *Eco*RI or *Bam*HI and *Eco*RI (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). The pET28b, pET30a, pGEX4T-3 and *H. pylori* DNA sequences were all electrophoresed on a 1% agarose gel and purified using the Wizard PCR Preps kit (Promega Corp., Madison WI, USA). Following ligation of the purified, digested expression vector and the purified, digested *H. pylori* DNA sequences, the products of the ligation reaction were transformed into *E. coli* JM109 competent cells (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). Individual bacterial colonies were screened for those containing the correct recombinant plasmids by preparing plasmid DNA as described above followed by analysis by restriction digestion profiles and DNA sequencing (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). These recombinant plasmids were then used to transform specific *E. coli* expression strains.

Transformation of competent bacteria with recombinant expression plasmids

Competent bacterial strains BL21(DE3), BL21(DE3)pLysS, HMS174(DE3) and HMS174(DE3)pLysS were prepared and transformed with the recombinant pET28b expression plasmids carrying the cloned *H. pylori* sequences according to standard methods (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). These expression host strains contain a chromosomal copy of the gene for T7 RNA polymerase. These hosts are lysogens of bacteriophage DE3, a lambda derivative that carries the *lacI* gene, the *lacUV5* promoter and the gene for T7

RNA polymerase. T7 RNA polymerase expression is induced by the addition of isopropyl-β-D-thiogalactoside (IPTG), and the T7 RNA polymerase then transcribes any taget plasmid, such as pET28b, that carries a T7 promoter sequence and a gene of interest.

Competent bacterial strains JM109 and DH5α were prepared and transformed with the recombinant pGEX4T-3 expression plasmid carrying the cloned *H. pylori* sequences according to standard methods (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994).

Expression of recombinant H. pylori sequences in E. coli

Transformants were collected from LB agar plates containing 25ug/ml kanamycin sulfate (ensures maintenance of the pET28b-based recombinant plasmids) or 100ug/ml ampicillin (ensures maintenance of the pGEX4T-3-based recombinant plasmids) and used to inoculate LB broth containing 25ug/ml kanamycin sulfate or 100ug/ml ampicillin and grown to an optical density at 600nm of 0.5 to 1.0 OD units, at which point 1mM IPTG was added to the culture for one to three hours to induce gene expression of the *H. pylori* recombinant DNA constructions. After induction of gene expression with IPTG, bacteria were pelleted by centrifugation and resuspended in SDS-PAGE solubilization buffer and subjected to SDS-PAGE (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). Proteins were visualized by staining with Coomassie Brilliant Blue or detected by western immunoblotting using the specific anti-His tag monoclonal antibody (Clontech, Palo Alto, CA, USA) or the anti-GST tag antibody (Pharmacia LKB) using standard methods (Current Protocols in Molecular Biology, John Wiley and Sons, Inc., F. Ausubel *et al.*, eds., 1994). The host strain that provided the highest level of recombinant protein production was then chosen for use in a large-scale induction in order to purify the recombinant protein. The strains used were HMS174(DE3) (pET28b-based constructs) and DH5α (pGEX4T-3-based constructs).

Removal of the C-terminal regions appeared in both systems to improve the level of expression, although this increase was far more prominent in the GST-fusion system. All recombinant proteins produced were of the predicted molecular weight based on the DNA sequence plus, if necessary, the size of the fusion tag. The truncated portion of the *H. pylori* protein contains some extremely hydrophobic stretches, and removal of these may be the reason for the increased expression.

EQUIVALENTS

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments and methods described herein. Such equivalents are intended to be encompassed by the scope of the following claims.

CLAIMS

1. An isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* polypeptide at least about 60% homologous to an amino acid sequence selected from the group consisting of SEQ ID NO: 98-SEQ ID NO: 194.

2. An isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* polypeptide selected from the group consisting of SEQ ID NO: 98-SEQ ID NO: 194.

3. An isolated nucleic acid which encodes an *H. pylori* polypeptide, comprising a nucleotide sequence at least about 60% homologous to a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 97, or a complement thereof.

4. The isolated nucleic acid of claim 1, comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 97, or a complement thereof.

5. An isolated nucleic acid molecule encoding an *H. pylori* polypeptide, comprising a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule comprising the nucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 97, or a complement thereof.

6. An isolated nucleic acid comprising a nucleotide sequence of at least 8 nucleotides in length, wherein the sequence hybridizes under stringent hybridization conditions to a nucleic acid having a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 97, or a complement thereof.

7. An isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* cell envelope polypeptide or a fragment thereof, said nucleic acid selected from the group consisting of SEQ ID NO: 63, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 61, SEQ ID

NO: 79, SEQ ID NO: 80, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 91, SEQ ID NO: 94, SEQ ID NO: 5, SEQ ID NO: 11, SEQ ID NO: 26, SEQ ID NO: 36, SEQ ID NO: 42, SEQ ID NO: 52, SEQ ID NO: 22, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 6, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 60, SEQ ID NO: 69, and SEQ ID NO: 83, or a complement thereof.

8. The isolated nucleic acid of claim 7, wherein said *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* flagella-associated polypeptide or a fragment thereof comrising a nucleotide sequence of SEQ ID NO: 63, or a complement thereof.

9. The isolated nucleic acid of claim 7, wherein said *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* inner membrane polypeptide or a fragment thereof encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 38, and SEQ ID NO: 39, or a complement thereof.

10. The isolated nucleic acid of claim 9, wherein said *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in transport encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 43, and SEQ ID NO: 44, or a complement thereof.

11. The isolated nucleic acid of claim 7, wherein said *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* outer membrane polypeptide or a fragment thereof encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 61, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 91, SEQ ID NO: 94, SEQ ID NO: 5, SEQ ID NO: 11, SEQ ID NO: 26, SEQ ID NO: 36, SEQ ID NO: 42, SEQ ID NO: 52, SEQ ID NO: 22, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 65, and SEQ ID NO: 66, or a complement thereof.

12. The isolated nucleic acid of claim 11, wherein said *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue or a fragment thereof encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 42, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 61, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 91, and SEQ ID NO: 94, or a complement thereof.

13. The isolated nucleic acid of claim 12, wherein said *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue and a C-terminal tyrosine cluster or a fragment thereof encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 11, SEQ ID NO: 26, SEQ ID NO: 36, SEQ ID NO: 42, and SEQ ID NO: 52, or a complement thereof.

14. An isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* cell envelope polypeptide or a fragment thereof, wherein said polypeptide is selected from the group consisting of SEQ ID NO: 160, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 158, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 188, SEQ ID NO: 191, SEQ ID NO: 102, SEQ ID NO: 108, SEQ ID NO: 123, SEQ ID NO: 133, SEQ ID NO: 139, SEQ ID NO: 149, SEQ ID NO: 119, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 162, SEQ ID NO: 163, SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 103, SEQ ID NO: 131, SEQ ID NO: 132, SEQ ID NO: 157, SEQ ID NO: 166, and SEQ ID NO: 180.

15. The isolated nucleic acid of claim 14, wherein said *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* flagella-associated polypeptide or a fragment thereof comprising an amino acid sequence of SEQ ID NO: 160.

16. The isolated nucleic acid of claim 14, wherein said *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* inner membrane polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 135, and SEQ ID NO: 136.

17. The isolated nucleic acid of claim 16, wherein said *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in transport selected from the group consisting of SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 140, and SEQ ID NO: 141.

18. The isolated nucleic acid of claim 14, wherein said *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* outer membrane polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 158, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 188, SEQ ID NO: 191, SEQ ID NO: 102, SEQ ID NO: 108, SEQ ID NO: 123, SEQ ID NO: 133, SEQ ID NO: 139, SEQ ID NO: 149, SEQ ID NO: 119, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 162, and SEQ ID NO: 163.

19. The isolated nucleic acid of claim 18, wherein said *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue or a fragment thereof selected from the group consisting of SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 108, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 123, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 133, SEQ ID NO: 139, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 149, SEQ ID NO: 158, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 188, and SEQ ID NO: 191.

20. The isolated nucleic acid of claim 19, wherein said *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue and a C-terminal tyrosine cluster or a fragment thereof

selected from the group consisting of SEQ ID NO: 108, SEQ ID NO: 123, SEQ ID NO: 133, SEQ ID NO: 139, and SEQ ID NO: 149.

21. An isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* cytoplasmic polypeptide or a fragment thereof, said nucleic acid selected from the group consisting of SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 92, and SEQ ID NO: 93, or a complement thereof.

22. The isolated nucleic acid of claim 21, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in mRNA translation, said nucleic acid selected from the group consisting of SEQ ID NO: 57 and SEQ ID NO: 58, or a complement thereof.

23. The isolated nucleic acid of claim 21, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in genome replication, transcription, recombination and repair, said nucleic acid selected from the group consisting of SEQ ID NO: 86 and SEQ ID NO: 87, or a complement thereof.

24. An isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* cytoplasmic polypeptide or a fragment thereof, wherein said polypeptide is selected from the group consisting of SEQ ID NO: 154, SEQ ID NO: 155, SEQ ID NO: 183, SEQ ID NO: 184, SEQ ID NO: 185, SEQ ID NO: 186, SEQ ID NO: 189, and SEQ ID NO: 190.

25. The isolated nucleic acid of claim 24, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in mRNA translation selected from the group consisting of SEQ ID NO: 154 and SEQ ID NO: 155.

26. The isolated nucleic acid of claim 24, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in genome replication, transcription, recombination and repair selected from the group consisting of SEQ ID NO: 183 and SEQ ID NO: 184.

27. An isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* secreted polypeptide or a fragment thereof, said nucleic acid selected from the group consisting of SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 20, SEQ ID NO: 25, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 53 SEQ ID NO: 64, SEQ ID NO: 67, SEQ ID NO: 70, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 81, SEQ ID NO: 82, SEQ ID NO: 90, SEQ ID NO: 95, and SEQ ID NO: 97, or a complement thereof.

28. An isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* secreted polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 107, SEQ ID NO: 109, SEQ ID NO: 117, SEQ ID NO: 122, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 142, SEQ ID NO: 143, SEQ ID NO: 150 SEQ ID NO: 161, SEQ ID NO: 164, SEQ ID NO: 167, SEQ ID NO: 174, SEQ ID NO: 175, SEQ ID NO: 178, SEQ ID NO: 179, SEQ ID NO: 187, SEQ ID NO: 192, and SEQ ID NO: 194.

29. An isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* cellular polypeptide or a fragment thereof, said nucleic acid selected from the group consisting of SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 21, SEQ ID NO: 33, SEQ ID NO: 37, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 47, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56 SEQ ID NO: 59, SEQ ID NO: 62, SEQ ID NO: 68, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, and SEQ ID NO: 96, or a complement thereof.

30. An isolated nucleic acid comprising a nucleotide sequence encoding an *H. pylori* cellular polypeptide or a fragment thereof, wherein said polypeptide is selected from the group consisting of SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 118, SEQ ID NO: 130, SEQ ID NO: 134, SEQ ID NO: 137, SEQ ID NO: 138, SEQ ID NO: 144, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 153, SEQ ID NO: 156, SEQ ID NO: 159, SEQ ID NO: 165, SEQ ID NO: 168, SEQ ID NO: 169, SEQ ID NO: 170, SEQ ID NO: 171, SEQ ID NO: 172, SEQ ID NO: 173, and SEQ ID NO: 193.

31. A probe comprising a nucleotide sequence consisting of at least 8 nucleotides of a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 97, or a complement thereof.

32. A recombinant expression vector comprising the nucleic acid of any of claims 1, 2, 3, 4, 5, 6, 7, 14, 21, 24, 27, 28, 29 or 30 operably linked to a transcription regulatory element.

33. A cell comprising a recombinant expression vector of claim 32.

34. A method for producing an *H. pylori* polypeptide comprising culturing a cell of claim 33 under conditions that permit expression of the polypeptide.

35. The method of claim 34, further comprising purifying the polypeptide from the cell.

36. A method for detecting the presence of a *Helicobacter* nucleic acid in a sample comprising:

(a) contacting a sample with a nucleic acid of any of claims 6 or 31 so that a hybrid can form between the probe and a *Helicobacter* nucleic acid in the sample; and

(b) detecting the hybrid formed in step (a), wherein detection of a hybrid indicates the presence of a *Helicobacter* nucleic acid in the sample.

37. An isolated *H. pylori* polypeptide comprising an amino acid sequence at least about 60% homologous to an *H. pylori* polypeptide selected from the group consisting of SEQ ID NO: 98-SEQ ID NO: 194.

38. An isolated *H. pylori* polypeptide which is encoded by a nucleic acid comprising a nucleotide sequence at least about 60% homologous to a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 97.

39. The isolated *H. pylori* polypeptide of claim 28, wherein said polypeptide is encoded by a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 97.

40. An isolated *H. pylori* polypeptide which is encoded by a nucleic acid which hybridizes under stringent hybridization conditions to a nucleic acid selected from the group consisting of SEQ ID NO: 1-SEQ ID NO: 97, or a complement thereof.

41. An isolated *H. pylori* polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 97-SEQ ID NO: 194.

42. An isolated *H. pylori* cell envelope polypeptide or a fragment thereof, wherein said polypeptide is selected from the group consisting of SEQ ID NO: 160, SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 158, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 188, SEQ ID NO: 191, SEQ ID NO: 102, SEQ ID NO: 108, SEQ ID NO: 123, SEQ ID NO: 133, SEQ ID NO: 139, SEQ ID NO: 149, SEQ ID NO: 119, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 162, SEQ ID NO: 163, SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 135, SEQ ID NO: 136, SEQ ID NO: 98, SEQ ID NO: 99, SEQ ID NO: 103, SEQ ID NO: 131, SEQ ID NO: 132, SEQ ID NO: 157, SEQ ID NO: 166, and SEQ ID NO: 180.

43. The isolated polypeptide of claim 42, wherein said *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* flagella-associated polypeptide or a fragment thereof comprising an amino acid sequence of SEQ ID NO:160.

44. The isolated polypeptide of claim 43, wherein said *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* inner membrane polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 140, SEQ ID NO: 141, SEQ ID NO: 135, SEQ ID NO: 136.

45. The isolated polypeptide of claim 44, wherein said *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in transport selected from the group consisting of SEQ ID NO: 145, SEQ ID NO: 146, SEQ ID NO: 114, SEQ ID NO: 115, SEQ ID NO: 116, SEQ ID NO: 140, and SEQ ID NO: 141.

46. The isolated polypeptide of claim 43, wherein said *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* outer membrane polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 104, SEQ ID

NO: 105, SEQ ID NO: 106, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 158, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 188, SEQ ID NO: 191, SEQ ID NO: 102, SEQ ID NO: 108, SEQ ID NO: 123, SEQ ID NO: 133, SEQ ID NO: 139, SEQ ID NO: 149, SEQ ID NO: 119, SEQ ID NO: 126, SEQ ID NO: 127, SEQ ID NO: 162, and SEQ ID NO: 163.

47. The isolated polypeptide of claim 46, wherein said *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue or a fragment thereof selected from the group consisting of SEQ ID NO: 104, SEQ ID NO: 105, SEQ ID NO: 106, SEQ ID NO: 108, SEQ ID NO: 110, SEQ ID NO: 111, SEQ ID NO: 120, SEQ ID NO: 121, SEQ ID NO: 123, SEQ ID NO: 124, SEQ ID NO: 125, SEQ ID NO: 133, SEQ ID NO: 139, SEQ ID NO: 147, SEQ ID NO: 148, SEQ ID NO: 149, SEQ ID NO: 158, SEQ ID NO: 176, SEQ ID NO: 177, SEQ ID NO: 181, SEQ ID NO: 182, SEQ ID NO: 188, and SEQ ID NO: 191.

48. The isolated polypeptide of claim 47, wherein said *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue and a C-terminal tyrosine cluster or a fragment thereof selected from the group consisting of SEQ ID NO: 108, SEQ ID NO: 123, SEQ ID NO: 133, SEQ ID NO: 139, and SEQ ID NO: 149.

49. An isolated *H. pylori* cell envelope polypeptide or a fragment thereof, wherein said polypeptide is encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 63, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 61, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 91, SEQ ID NO: 94, SEQ ID NO: 5, SEQ ID NO: 11, SEQ ID NO: 26, SEQ ID NO: 36, SEQ ID NO: 42, SEQ ID NO: 52, SEQ ID NO: 22, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 65, SEQ ID NO: 66, SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 38, SEQ ID NO: 39, SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 6, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 60, SEQ ID NO: 69, and SEQ ID NO: 83.

50. The isolated polypeptide of claim 49, wherein said *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* flagella-associated polypeptide or a fragment thereof encoded by a nucleic acid comprising a nucleotide sequence of SEQ ID NO: 63.

51. The isolated polypeptide of claim 49, wherein said *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* inner membrane polypeptide or a fragment thereof encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 43, SEQ ID NO: 44, SEQ ID NO: 38, SEQ ID NO: 39.

52. The isolated polypeptide of claim 51, wherein said *H. pylori* inner membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in transport encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 48, SEQ ID NO: 49, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 43, and SEQ ID NO: 44.

53. The isolated polypeptide of claim 49, wherein said *H. pylori* cell envelope polypeptide or a fragment thereof is an *H. pylori* outer membrane polypeptide or a fragment thereof encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 61, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 91, SEQ ID NO: 94, SEQ ID NO: 5, SEQ ID NO: 11, SEQ ID NO: 26, SEQ ID NO: 36, SEQ ID NO: 42, SEQ ID NO: 52, SEQ ID NO: 22, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 65, and SEQ ID NO: 66.

54. The isolated polypeptide of claim 53, wherein said *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue or a fragment thereof encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 36, SEQ ID NO: 42, SEQ ID NO: 50, SEQ ID NO: 51, SEQ ID NO: 52, SEQ ID NO: 61, SEQ ID NO: 79, SEQ ID NO: 80, SEQ ID NO: 84, SEQ ID NO: 85, SEQ ID NO: 91, and SEQ ID NO: 94.

55. The isolated polypeptide of claim 54, wherein said *H. pylori* outer membrane polypeptide or a fragment thereof is an *H. pylori* polypeptide having a terminal phenylalanine residue and a C-terminal tyrosine cluster or a fragment thereof encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 11, SEQ ID NO: 26, SEQ ID NO: 36, SEQ ID NO: 42, and SEQ ID NO: 52.

56. An isolated *H. pylori* cytoplasmic polypeptide or a fragment thereof, wherein said polypeptide is selected from the group consisting of SEQ ID NO: 154, SEQ ID NO: 155, SEQ ID NO: 183, SEQ ID NO: 184, SEQ ID NO: 185, SEQ ID NO: 186, SEQ ID NO: 189, and SEQ ID NO: 190.

57. The isolated polypeptide of claim 56, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in mRNA translation selected from the group consisting of SEQ ID NO: 154 and SEQ ID NO: 155.

58. The isolated polypeptide of claim 56, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in genome replication, transcription, recombination and repair selected from the group consisting of SEQ ID NO: 183 and SEQ ID NO: 184.

59. An isolated *H. pylori* cytoplasmic polypeptide or a fragment thereof, wherein said polypeptide is encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 57, SEQ ID NO: 58, SEQ ID NO: 86, SEQ ID NO: 87, SEQ ID NO: 88, SEQ ID NO: 89, SEQ ID NO: 92, and SEQ ID NO: 93.

60. The isolated polypeptide of claim 59, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in mRNA translation, said polypeptide encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 57 and SEQ ID NO: 58.

61. The isolated polypeptide of claim 59, wherein said *H. pylori* cytoplasmic polypeptide or a fragment thereof is an *H. pylori* polypeptide or a fragment thereof involved in genome replication, transcription, recombination and repair, said polypeptide encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 86 and SEQ ID NO: 87.

62. An isolated *H. pylori* cellular polypeptide or a fragment thereof, wherein said polypeptide is selected from the group consisting of SEQ ID NO: 112, SEQ ID NO: 113, SEQ ID NO: 118, SEQ ID NO: 130, SEQ ID NO: 134, SEQ ID NO: 137, SEQ ID NO: 138, SEQ ID NO: 144, SEQ ID NO: 151, SEQ ID NO: 152, SEQ ID NO: 153, SEQ ID NO: 156, SEQ ID NO: 159, SEQ ID NO: 165, SEQ ID NO: 168, SEQ ID NO: 169, SEQ ID NO: 170, SEQ ID NO: 171, SEQ ID NO: 172, SEQ ID NO: 173, and SEQ ID NO: 193.

63. An isolated *H. pylori* cellular polypeptide or a fragment thereof, wherein said polypeptide is encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 21, SEQ ID NO: 33, SEQ ID NO: 37, SEQ ID NO: 40, SEQ ID NO: 41, SEQ ID NO: 47, SEQ ID NO: 54, SEQ ID NO: 55, SEQ ID NO: 56 SEQ ID NO: 59, SEQ ID NO: 62, SEQ ID NO: 68, SEQ ID NO: 71, SEQ ID NO: 72, SEQ ID NO: 73, SEQ ID NO: 74, SEQ ID NO: 75, SEQ ID NO: 76, and SEQ ID NO: 96.

64. An isolated *H. pylori* secreted polypeptide or a fragment thereof, wherein said polypeptide is selected from the group consisting of SEQ ID NO: 100, SEQ ID NO: 101, SEQ ID NO: 107, SEQ ID NO: 109, SEQ ID NO: 117, SEQ ID NO: 122, SEQ ID NO: 128, SEQ ID NO: 129, SEQ ID NO: 142, SEQ ID NO: 143, SEQ ID NO: 150 SEQ ID NO: 161, SEQ ID NO: 164, SEQ ID NO: 167, SEQ ID NO: 174, SEQ ID NO: 175, SEQ ID NO: 178, SEQ ID NO: 179, SEQ ID NO: 187, SEQ ID NO: 192, and SEQ ID NO: 194.

65. An isolated *H. pylori* secreted polypeptide or a fragment thereof, wherein said polypeptide is encoded by a nucleic acid selected from the group consisting of SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 10, SEQ ID NO: 12, SEQ ID NO: 20, SEQ ID NO: 25, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 45, SEQ ID NO: 46, SEQ ID NO: 53 SEQ ID NO: 64, SEQ ID NO: 67, SEQ ID NO: 70, SEQ ID NO: 77, SEQ ID NO: 78, SEQ ID NO: 81, SEQ ID NO: 82, SEQ ID NO: 90, SEQ ID NO: 95, and SEQ ID NO: 97.

66. A fusion protein comprising an *H. pylori* polypeptide which comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 98-SEQ ID NO: 194 operatively linked to a non-*H. pylori* polypeptide.

67. A vaccine formulation for prophylactic or therapeutic treatment of an *H. pylori* infection comprising an effective amount of at least one isolated nucleic acid of any of claims 1, 2, 3, 4, 5, 6, 7, 14, 21, 24, 27, 28, 29 or 30.

68. A vaccine formulation for prophylactic or therapeutic treatment of an *H. pylori* infection comprising an effective amount of at least one *H. pylori* polypeptide or a fragment thereof of any of claims 37, 38, 40, 41, 42, 49, 56, 59, 62, 63, 64 or 65.

69. A vaccine formulation of claim 67, further comprising a pharmaceutically acceptable carrier.

70. A vaccine formulation of claim 68, further comprising a pharmaceutically acceptable carrier.

71. A vaccine formulation of claim 69, wherein the pharmaceutically acceptable carrier comprises an adjuvant.

72. A vaccine formulation of claim 70, wherein the pharmaceutically acceptable carrier comprises an adjuvant.

73. A vaccine formulation of claim 69, wherein the pharmaceutically acceptable carrier comprises a delivery system.

74. A vaccine formulation of claim 70, wherein the pharmaceutically acceptable carrier comprises a delivery system.

75. A vaccine formulation of claim 73, wherein the delivery system comprises a live vector.

76. A vaccine formulation of claim 74, wherein the delivery system comprises a live vector.

77. A vaccine formulation of claim 75, wherein the live vector is a bacteria or a virus.

78. A vaccine formulation of claim 76, wherein the live vector is a bacteria or a virus.

79. A vaccine formulation of claim 73, wherein the pharmaceutically acceptable carrier further comprises an adjuvant.

80. A vaccine formulation of claim 74, wherein the pharmaceutically acceptable carrier further comprises an adjuvant.

81. A vaccine formulation for prophylactic or therapeutic treatment of an *H. pylori* infection comprising an effective amount of at least one isolated nucleic acid encoding an *H. pylori* outer membrane polypeptide or a fragment thereof, said nucleic acid selected from the group consisting of SEQ ID NO: 28, SEQ ID NO: 50, SEQ ID NO: 24, SEQ ID NO: 11, SEQ ID NO: 52, SEQ ID NO: 42 and SEQ ID NO: 79.

82. The vaccine formulation of claim 81, wherein said nucleic acid comprises a nucleotide sequence of SEQ ID NO: 52.

83. A vaccine formulation for prophylactic or therapeutic treatment of an *H. pylori* infection comprising an effective amount of at least one *H. pylori* outer membrane polypeptide or a fragment thereof, said polypeptide selected from the group consisting of SEQ ID NO: 125, SEQ ID NO: 147, SEQ ID NO: 121, SEQ ID NO: 108, SEQ ID NO: 149, SEQ ID NO: 139 and SEQ ID NO: 176.

84. The vaccine formulation of claim 81, wherein said polypeptide comprises an amino acid sequence of SEQ ID NO: 149.

85. A vaccine formulation of claims 81 or 83, further comprising a pharmaceutically acceptable carrier.

86. A vaccine formulation of claim 85, wherein the pharmaceutically acceptable carrier comprises an adjuvant.

87. A vaccine formulation of claim 85, wherein the pharmaceutically acceptable carrier comprises a delivery system.

88. A vaccine formulation of claim 87, wherein the delivery system comprises a live vector.

89. A vaccine formulation of claim 88, wherein the live vector is a bacteria or a virus.

90. A vaccine formulation of claim 86, wherein the pharmaceutically acceptable carrier further comprises an adjuvant.

91. A method of treating or reducing a risk of *H. pylori* infection in a subject comprising administering to a subject a vaccine formulation of claim 67, such that treatment or reduction of risk of *H. pylori* infection occurs.

92. A method of treating or reducing a risk of *H. pylori* infection in a subject comprising administering to a subject a vaccine formulation of claim 68, such that treatment or reduction of risk of *H. pylori* infection occurs.

93. A method of treating or reducing a risk of *H. pylori* infection in a subject comprising administering to a subject a vaccine formulation of claim 81, such that treatment or reduction of risk of *H. pylori* infection occurs.

94. A method of treating or reducing a risk of *H. pylori* infection in a subject comprising administering to a subject a vaccine formulation of claim 83, such that treatment or reduction of risk of *H. pylori* infection occurs.

95. A method of producing a vaccine formulation comprising: combining at least one isolated *H. pylori* polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 98-SEQ ID NO: 194 with a pharmaceutically acceptable carrier to thereby form a vaccine formulation.

96. A method of producing a vaccine formulation comprising:

(a) providing at least one isolated *H. pylori* polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 98-SEQ ID NO: 194; and

(b) combining at least one said isolated *H. pylori* polypeptide or a fragment thereof with a pharmaceutically acceptable carrier to thereby form a vaccine formulation.

97. A method of producing a vaccine formulation comprising:

(a) culturing a cell under condition that permit expression of an *H. pylori* polypeptide or a fragment thereof selected from the group consisting of SEQ ID NO: 98-SEQ ID NO: 194;

(b) isolating said *H. pylori* polypeptide or a fragment thereof from said cell; and

(c) combining at least one said isolated *H. pylori* polypeptide or a fragment thereof with a pharmaceutically acceptable carrier to thereby form a vaccine formulation.

98. A chimeric *H. pylori* polypeptide comprising at least two *H. pylori* polypeptides or fragments thereof, wherein said polypeptides are encoded by nucleic acid sequences selected from the group consisting of SEQ ID NO:1-SEQ ID NO:97.

99. A chimeric *H. pylori* polypeptide comprising at least two *H. pylori* polypeptides or fragments thereof, wherein said polypeptides are selected from the group consisting of SEQ ID NO:98-SEQ ID NO:194.

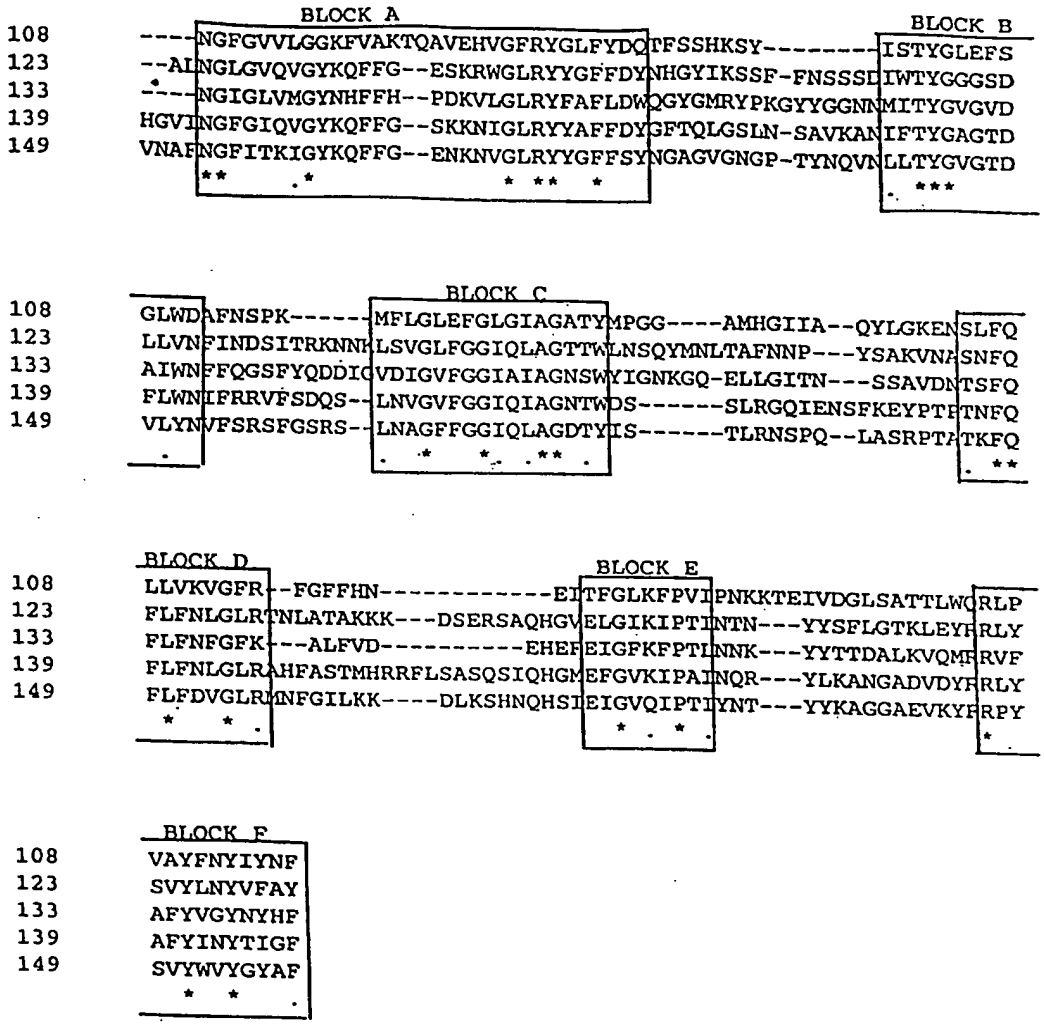

BLOCK A
BLOCK B
BLOCK C
BLOCK D
BLOCK E
BLOCK F
108 ----NGFGVVLGGKFVAKTQAVEHVGFRYGLFYDQTFSSHKSY--------ISTYGLEFS
123 --ALNGLGVQVGYKQFFG--ESKRWGLRYYGFFDYNHGYIKSSF-FNSSSDIWTYGGGSD
133 ----NGIGLVMGYNHFFH--PDKVLGLRYFAFLDWQGYGMRYPKGYYGGNNMITYGVGVD
139 HGVINGFGIQVGYKQFFG--SKKNIGLRYYAFFDYGFTQLGSLN-SAVKANIFTYGAGTD
149 VNAFNGFITKIGYKQFFG--ENKNVGLRYYGFFSYNGAGVGNGP-TYNQVNLLTYGVGTD
108 GLWDAFNSPK------MFLGLEFGLGIAGATYMPGG----AMHGIIA--QYLGKENSLFQ
123 LLVNFINDSITRKNNKLSVGLFGGIQLAGTTWLNSQYMNLTAFNNP---YSAKVNASNFQ
133 AIWNFFQGSFYQDDICVDIGVFGGIAIAGNSWYIGNKGQ-ELLGITN---SSAVDNTSFQ
139 FLWNIFRRVFSDQS--LNVGVFGGIQIAGNTWDS------SLRGQIENSFKEYPTPTNFQ
149 VLYNVFSRSFGSRS--LNAGFFGGIQLAGDTYIS------TLRNSPQ--LASRPTATKFQ
108 LLVKVGFR--FGFFHN----------EITFGLKFPVIPNKKTEIVDGLSATTLWQRLP
123 FLFNLGLRTNLATAKKK---DSERSAQHGVELGIKIPTINTN---YYSFLGTKLEYRRLY
133 FLFNFGFK---ALFVD----------EHEFEIGFKFPTLNNK---YYTTDALKVQMRRVF
139 FLFNLGLRAHFASTMHRRFLSASQSIQHGMEFGVKIPAINQR---YLKANGADVDYRRLY
149 FLFDVGLRMNFGILKK----DLKSHNQHSIEIGVQIPTIYNT---YYKAGGAEVKYRRPY
108 VAYFNYIYNF
123 SVYLNYVFAY
133 AFYVGYNYHF
139 AFYINYTIGF
149 SVYWVYGYAF

Figure 1

```
139     MKNTNTKEIKNTRMKKGYSQYHTLKKGLLKTALLFSLPLSVALAEDDGFYMGVGYQIGGA
149     -MKQNLKPFK--MIKENLMTQSQKVRFLAPLSLALSLSFNPVGAEEDGGFMTFGYELGQV
102     ------------MKKKFLSLT-----------LGSLLVSALSAEDNGFFVSAGYQIGES
                  .*                    **       **. * ..  **..*

139     QQNINNKG----------STLRNNVIDDFRQVGVGMAGGNG---------LLALATNT
149     VQQVKNPGKIKAEELAGLLNSTTTNNTNINIAGTGGNVAGTLGNLFMNQLGNLIDLYPTL
102     AQMVKNTKG----------IQDLSDSYERLNNLLTNYSV---------LNALIRQS
         * . *                    .       .  ...           *  *
```

Figure 2

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/22104

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : A61K 39/00; C12Q 1/68; A01N 43/04
US CL : 514/44; 435/6; 424/185.1

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 514/44; 435/6; 424/185.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

genbank
search terms: specific amino acid and nucleic acid sequences

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	HIRSCHL, A M et al. Comparison of ELISA antigen preparation alone or in combination for serodiagnosing Helicobacter pylori infection. 1990, Vol. 43, pages 511-513.	1-99

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 30 MARCH 1998	Date of mailing of the international search report 21 APR 1998
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer GINNY PORTNER Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/22104

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
3 groups of 4 sequences: SEQ ID 5, 11, 14, 26, 28, 36, 42, 51, 52, 79, 102,108,111,121,123,125,133,139,148,149,176,177

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest
☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/22104

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s)1-34, 36, 67, 69,71,73,75,77, 79, 81-82, 85-91, 93, 98, drawn to 178 isolated nucleic acids, vectors containing the different nucleic acids, organisms transformed with the DNA's, vaccines and method of producing the polypeptides encoded by the no fewer than 178 different DNA's.

Group II, claim(s) 35, 37-66, 68, 70, 72, 74, 76, 78, 80, 83-90, 92, 94-97, 99, drawn to 97 isolated polypeptides encoded by a the DNA's mentioned in Group I.

This application contains claims directed to more than one species of the generic invention. These species are deemed to lack Unity of Invention because they are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for more than one species to be searched, the appropriate additional search fees must be paid. The species are as follows:

Group I contains a separate DNA species for each sequence mentioned. Therefore, theis a minimun of 178 species. Group II contains one polypetide for each amino acid seuqence mentioned; therefore, there is a minimum of 68 species in Group II.

For either Group that Applicant elects, a total of 10 (ten) specified sequences will be searched and no more than 4 (four) specified sequences will be searched for each additional fee paid; if no addional fee is paid and no election indicated the first 10 sequences will be searched.

The inventions listed as Groups I and II do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: The polypeptide encoding DNAs, vectors containing them, organisms tranformed with them and methods of polypeptide production using them are materially different from each other and are therefore independent form the polypeptides of Group II. Additionally, none of the products or methods of Group I is needed to make the polypeptides of Group II.

The species listed above do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, the species lack the same or corresponding special technical features for the following reasons: There is no relationship between or among the various nucleotide and amino acid sequences mentioned in the claims.

Applicant elected the first 10 sequences to be, SEQ ID NO. 5,11,14,26,28,36,42,51,52 and 79.